

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2001, 10:44:24 ; Search time 379.17 Seconds
(without alignments)
8147.469 Million cell updates/sec

Title: US-09-269-874-1

Perfect score: 4920

Sequence: 1 atgaagatcatatctttt.....tattatcacgttcaattaa 4920

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq.0601:*

- 1: /SIDSI/gcgdata/geneseq/geneseqn/NA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseqn/NA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseqn/NA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseqn/NA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseqn/NA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseqn/NA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseqn/NA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseqn/NA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseqn/NA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseqn/NA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseqn/NA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseqn/NA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseqn/NA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseqn/NA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseqn/NA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseqn/NA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseqn/NA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseqn/NA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4900.8	99.6	4940	19	AAV35363
2	4779.2	97.1	5760	6	AAV50530
3	2812.8	57.2	4940	19	AAV21451
4	2694.8	54.8	5181	16	AAO80911
5	1798.8	36.6	1950	18	AAV97956
6	1785.4	36.3	1897	18	AAV97958
7	1741.2	35.4	1896	18	AAV97957
8	1063.4	21.6	1088	20	AAV56009
9	1063.4	13.0	1077	22	AAV25587
10	641.4	13.0	1077	22	AAV68976
11	601	12.2	1065	20	AAV56008

12	601	12.2	1065	20	AAV25586	Merozoite surface
13	588	12.0	1140	20	AAV56021	Merozoite surface
14	588	12.0	1140	20	AAV25593	Merozoite surface
15	535	10.9	1068	9	AAV81151	DNA sequence encod
16	525.8	10.7	660	9	AAV81148	DNA sequence encod
17	448.2	9.1	737	6	AAV50355	Storage-specific.
18	439.4	8.9	786	22	AAV68978	Merozoite surface
19	435.2	8.8	668	9	AAV82176	31-1 Repeated Dete
20	404.2	8.2	462	9	AAV81149	DNA sequence encod
21	402.6	8.2	456	9	AAV81150	DNA sequence encod
22	282.6	5.7	936	22	AAV58252	Oligonucleotide D1
23	282.6	5.7	936	22	AAV58254	Oligonucleotide D1
24	282.6	5.7	936	22	AAV58257	Oligonucleotide D1
25	282.6	5.7	936	22	AAV58259	Oligonucleotide D2
26	282.6	5.7	936	22	AAV58262	Oligonucleotide D2
27	282.6	5.7	936	22	AAV58255	Oligonucleotide D1
28	280.4	5.7	936	22	AAV58252	Oligonucleotide D1
29	280.4	5.7	936	22	AAV58254	Oligonucleotide D1
30	280.4	5.7	936	22	AAV58257	Oligonucleotide D1
31	280.4	5.7	936	22	AAV58259	Oligonucleotide D2
32	280.4	5.7	936	22	AAV58262	Oligonucleotide D2
33	280.4	5.7	936	22	AAV58255	Oligonucleotide D1
34	243.4	4.9	5940	21	AAV70105	Plasmodium falci
35	240.2	4.9	306	6	AAV50354	Storage-specific.
36	232.8	4.7	354	18	AAV80403	PfMSPI (p19) A codin
37	232.8	4.7	354	18	AAV80450	PfMSPI (p19) A codin
38	207.2	4.2	333	22	AAV68977	Merozoite surface
39	188.4	3.8	11922	21	AAV70187	Plasmodium falci
40	185.4	3.8	387	18	AAV80404	PfMSPI (p19) S codin
41	185.4	3.8	387	18	AAV94549	PfMSPI (p19) S codin
42	181.8	3.7	3579	21	AAV70099	Plasmodium falci
43	176.8	3.6	5409	21	AAV70151	Plasmodium falci
44	176.8	3.6	6033	21	AAV70152	Plasmodium falci
45	176.4	3.6	7458	21	AAV70106	Plasmodium falci

ALIGNMENTS

RESULT 1	
ID AAV35363	AAV35363 standard; DNA; 4940 BP.
AC AAV35363;	
XX	
DT 23-SEP-1998	(first entry)
DE	
XX	
KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;	
KW monoclonal antibody; passive immunisation; parasite; ss.	
XX	
OS Plasmodium falciparum.	
XX	
FH Key	Location/Qualifiers
FT CDS	10..4929
FT	/tag= a
FT	/product= gp190
XX	
PN	
PD	MO9814583-A2.
XX	
PD	09-APR-1998.
XX	
PF	02-OCT-1997; 97WO-EP05441.
XX	
PR	02-OCT-1996; 96DE-4040817.
XX	
PA	(BUA/)
XX	
PI	Bujard H, Pan W, Tolle R;
XX	
DR	WPI: 1998-240088/21.
DR	P-PSDB; AAV54145.

XX Recombinant production of complete gp190/MSP-1 Plasmodium surface
PT protein - useful in anti-malaria vaccines, also stabilising genes by
PT reducing their AT content
XX
PS Example 1: Fig 3c; 48pp; German.
XX
CC This sequence encodes the Plasmodium falciparum gp190/MSP-1 (merozoite
CC surface protein). This gene is used in a method for stabilising the
CC gene sequences by reducing the AT content. Such products are useful in
CC vaccines against malaria or for producing monoclonal antibodies (for
CC passive immunisation). The complete gp190 protein can now be produced
CC outside the parasite and has, at least over extended regions, the native
CC pattern of folding. Larger amounts of the protein can be produced
CC recombinantly than would be possible using the parasites as source.
XX
SO Sequence 4940 BP; 2196 A; 597 C; 687 G; 1460 T; 0 other:

Query Match 99.6%; Score 4900.8; DB 19; Length 4940;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 4908; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 atgaagatcatattcttcttgcgttcattcctttttttattataaatacacaatgtgta 60
DB 10 atgaagatcatattcttcttgcgttcattcctttttttattataaatacacaatgtgta 69
QY 61 accatataaagttaacagaactgtcaaaaactagaagcttagaagatgacgtttg 120
DB 70 acacataaagaatttccagaactgtcaaaaactagaagcttagaagatgacgtttg 129
QY 121 acaggtatagttatttccaaaagaaaatggtatataatgaagaaacagtgagaca 180
DB 130 acaggtatagttatttccaaaagaaaatggtatataatgaagaaacagtgagaca 189
QY 181 gctgttcaactagttacactgtgttcaaaaggttcagttccttcagggttcagggtgc 240
DB 190 gctgttcaactagttacactgtgttcaaaaggttcagttccttcagggttcagggtgc 249
QY 241 tcaagttcctcaggttgctcagttcagttcaggttgctcaggttgctcaggt 300
DB 250 tcaagttcctcaggttgctcagttcagttcaggttgctcaggttgctcaggt 309
QY 301 gcttcaagttggttcaagttatccacaagcgtacaaatccttcagataatcagaatgtca 360
DB 310 gcttcaagttggttcaagttatccacaagcgtacaaatccttcagataatcagaatgtca 369
QY 361 gatgtcaatccttaacgtgtatataaacacagagttacgaataactgttaactatacaa 420
DB 370 gatgtcaatccttaacgtgtatataaacacagagttacgaataactgttaactatacaa 429
QY 421 gaactcaaatcctcctcaactccttgatttaactaatcatatgttaactgttgatcaat 480
DB 430 gaactcaaatcctcctcaactccttgatttaactaatcatatgttaactgttgatcaat 489
QY 481 attcatgtttcaataatataatgatgagatalgaagaaaataatgaatattatataaa 540
DB 490 attcatgtttcaataatataatgatgagatalgaagaaaataatgaatattatataaa 549
QY 541 ttaaacctttatttggatttaataagaagcaaaatataatgataatgataatgataat 600
DB 550 ttaaacctttatttggatttaataagaagcaaaatataatgataatgataatgataat 609
QY 601 tgcataatccttcaatcttaaatcgtgcaaatgaatagaagcttaataaaact 660
DB 610 tgcataatccttcaatcttaaatcgtgcaaatgaatagaagcttaataaaact 669
QY 661 gtttgcgatatagaacacattagacaatattaaagataatgagaagaaatggaagat 720
DB 670 gtttgcgatatagaacacattagacaatattaaagataatgagaagaaatggaagat 729
QY 721 tacatataaaaaataaaaaaacattagaataatataatgaatgaatgaagaagtaag 780
DB 730 tacatataaaaaataaaaaaacattagaataatataatgaatgaatgaagaagtaag 789

DB 730 tacatataaaaaataaaaaaacattagaataatataatgaatgaatgaagaagtaag 789
QY 781 aaaaactattataaaaataagaatgacacttaagaagaataaaaaaataattatacaa 840
DB 790 aaaaacttgataaaaataagaatgacacttaagaagaataaaaaaataattatacaa 849
QY 841 gctcaatgatccttcttatttcaatacaaatataagaagaacacataatataaagc 900
DB 850 gctcaatgatccttcttatttcaatacaaatataagaagaacacataatataaagc 909
QY 901 gttttagaaaacgatttgacactttaaaaaaataatgaacattagaagattactgtat 960
DB 910 gttttagaaaacgatttgacactttaaaaaaataatgaacattagaagattactgtat 969
QY 961 aagtaataataatataaaatcccccacggcaattctggaataatcaccaatctctc 1020
DB 970 aagtaataataatataaaatcccccacggcaattctggaataatcaccaatctctc 1029
QY 1021 ctgtataagacaaaataatcgaagacacgaaaagaataaagaatctgcaaaact 1080
DB 1030 ctgtataagacaaaataatcgaagacacgaaaagaataaagaatctgcaaaact 1089
QY 1081 attaaattatattgtagttatttactgactccactggaattagaatactattaga 1140
DB 1090 attaaattatattgtagttatttactgactccactggaattagaatactattaga 1149
QY 1141 gaaaataataaataattgtagtaagtgcaaaagttgaacaaagaaataactgaaacc 1200
DB 1150 gaaaataataaataattgtagtaagtgcaaaagttgaacaaagaaataactgaaacc 1209
QY 1201 aatgaatattccaatggagttacttactcctttgtcatalaagatatcaaatgtctta 1260
DB 1210 aatgaatattccaatggagttacttactcctttgtcatalaagatatcaaatgtctta 1269
QY 1261 aatgaactaatccttcttggtgtatataatcactgtatataacaaagaacagat 1320
DB 1270 aatgaactaatccttcttggtgtatataatcactgtatataacaaagaacagat 1329
QY 1321 aaaaactatatactgtagtaagaagaataaattataatgaatgaagaataaatt 1380
DB 1330 aaaaactatatactgtagtaagaagaataaattataatgaatgaagaataaatt 1389
QY 1381 aaaaataaaaaaaaataatgtatcgtataaaaaactcttaccgaagaacagatcagttc 1440
DB 1390 aaaaataaaaaaaaataatgtatcgtataaaaaactcttaccgaagaacagatcagttc 1449
QY 1441 ttaaatgatatcaaaaagaatatgaaaaacttaacttaatgaatattatgaagaacttc 1500
DB 1450 ttaaatgatatcaaaaagaatatgaaaaacttaacttaatgaatattatgaagaacttc 1509
QY 1501 aatataataatagatttaacttaatttcgaaaataatgatgggttaagaagattcataaa 1560
DB 1510 aatataataatagatttaacttaatttcgaaaataatgatgggttaagaagattcataaa 1569
QY 1561 gttgagaacattacacacataaactttgtgcatccatagtgaataatcttaacaatactt 1620
DB 1570 gttgagaacattacacacataaactttgtgcatccatagtgaataatcttaacaatactt 1629
QY 1621 gaaaagttaacaaaagctccttaatatatggaagattatctttaaaggaatagtagtt 1680
DB 1630 gaaaagttaacaaaagctccttaatatatggaagattatctttaaaggaatagtagtt 1689
QY 1681 gaaaagaattataaattataaataattataagaagaataagaagaatagatgtgaaca 1740
DB 1690 gaaaagaattataaattataaataattataagaagaataagaagaatagatgtgaaca 1749
QY 1741 ttagtgaataatataaagaatgagaagacgttttgaaaaaaaataattactaaagc 1800
DB 1750 ttagtgaataatataaagaatgagaagacgttttgaaaaaaaataattactaaagc 1809
QY 1801 gaaaataaacagatgaaaaaattttagaagttactgacattgttaaaatgacaagtcca 1860
DB 1810 gaaaataaacagatgaaaaaattttagaagttactgacattgttaaaatgacaagtcca 1869

QY	1861	aaagtattataatgaccaaaattgacgaattaaaaagactcaattgatttaaaaat	19270
Db	1870	aaagttttatttaagaaacaaattgacgaattaaaaaagactcaattgatttaaaaat	19293
QY	1921	gtagatttaaaaacataatacagtttcccaattcttcaaaaagaanaataagaagaa	19680
Db	1930	gtagaattaaaaacataatacagtttcccaattcttcaaaaagaanaataagaagaa	19699
QY	1981	ccttatataatcttgcttggaagaaagaanaattgataaattaaagtttcaatgctcag	20408
Db	1990	ccttatataatcttgcttggaagaaagaanaattgataaattaaagtttcaatgctcag	20493
QY	2041	gtagaatcatgtataaaatgaaagaaaaaaacataaaaacagaagttcaatcggataat	21000
Db	2050	gtagaatcatgtataaaatgaaagaaaaaaacataaaaacagaagttcaatcggataat	21099
QY	2101	tcggaacccattaacccggaaggagaataatacagacaaagcaactcaaaaactcggacaca	21600
Db	2110	tcggaacccattaacccggaaggagaataatacagacaaagcaactcaaaaactcggacaca	21699
QY	2161	gcgaatctgtcttagaaggagagattcaagtaacaagcccaagccaagaacaaaacaaaga	22200
Db	2170	gcgaagatcgtcttagaaggagagattcaagtaacaagcccaagccaagaacaaaacaaaga	22299
QY	2221	caaccaccagttaccagtaaccagtaaccagaagcaaaagccaagttcccaaccaccagca	22800
Db	2230	caaccaccagttaccagtaaccagtaaccagaagcaaaagccaagttcccaaccaccagca	22899
QY	2281	ccagtaataataaaactgaaaaatgtttcccaattagattacttgaaaaattatataaa	23400
Db	2290	ccagtaataataaaactgaaaaatgtttcccaattagattacttgaaaaattatataaa	23499
QY	2341	tttttaatacttcatataataitgcacaataatatttggtttcacctcaactatgaac	24000
Db	2350	tttttaatacttcatataataitgcacaataatatttggtttcacctcaactatgaac	24099
QY	2401	gaaagaatataaaacacataataattacaagaaggagaagaacaaattgaattcaagt	24600
Db	2410	gaaagaatataaaacacataataattacaagaaggagaagaacaaattgaattcaagt	24699
QY	2461	gaccatagagctatgtgttttaatalacaaaataaacatccctgtaagtattctatgttt	25200
Db	2470	gaccatagagctatgtgttttaatalacaaaataaacatccctgtaagtattctatgttt	25299
QY	2521	gatagtcttaaacaaatagtttatcacacaactttatggaattttagaagaaagaatgct	25800
Db	2530	gatagtcttaaacaaatagtttatcacacaactttatggaattttagaagaaagaatgct	25899
QY	2581	tgtaattatataaacactaaggataatgacaaattaaanaatttttagaggaagcgaaa	26400
Db	2590	tgtaattatataaacactaaggataatgacaaattaaanaatttttagaggaagcgaaa	26499
QY	2641	aaagatcacacatctgttaaaaactttcttaagtctcaatcaltgcaaccattitcatlaaa	27000
Db	2650	aaagatcacacatctgttaaaaactttcttaagtctcaatcaltgcaaccattitcatlaaa	27099
QY	2701	cctcaagataaaacccgaagtaagtgcgaataatgatatacatcattcaacaattggaat	27600
Db	2710	cctcaagataaaacccgaagtaagtgcgaataatgatatacatcattcaacaattggaat	27699
QY	2761	aatagtttaaanaattttggaacacatctgagctcttggaanaaanaaataataatacaca	28200
Db	2770	aatagtttaaanaattttggaacacatctgagctcttggaanaaanaaataataatacaca	28299
QY	2821	gaattataaggtcacaanaagtagtgaanaactttatgaaagaatataataagaattgagt	28800
Db	2830	gaattataaggtcacaanaagtagtgaanaactttatgaaagaatataataagaattgagt	28899
QY	2881	acattttataatgaaatccttttcaaaaattttgtcaaaaactcaaaagctgataattaatca	29400
Db	2890	acattttataatgaaatccttttcaaaaattttgtcaaaaactcaaaagctgataattaatca	29499

[illegible]

```

Db      4030 tcaagtaattatgtgtcgaagatccatataaattcttaataagaagaagaataaa 4089
         |||
Qy      4081 ttcttaagcgttatataatbattaagattcaatagatcggatataaatttgcaat 4140
         |||
Db      4090 ttcttaagcgttatataatbattaagattcaatagatcggatataaatttgcaat 4149
         |||
Qy      4141 gatgtcttgatataataaataataccgaataataatcagatttgatccaat 4200
         |||
Db      4150 gatgtcttgatataataaataataccgaataataatcagatttgatccaat 4209
         |||
Qy      4201 aaaaaatatcaacgacaacaaagtgaaatgagaataacccctccctttaacaat 4260
         |||
Db      4210 aaaaaatatcaacgacaacaaagtgaaatgagaataacccctccctttaacaat 4269
         |||
Qy      4261 attgagacctataataaagcgttaagataaattgattatttgatccaatccttgaa 4320
         |||
Db      4270 attgagacctataataaagcgttaagataaattgattatttgatccaatccttgaa 4329
         |||
Qy      4321 gcaaaaattctaaatatacatatgagaatcaaacgtagaagttaataataagaact 4380
         |||
Db      4330 gcaaaaattctaaatatacatatgagaatcaaacgtagaagttaataataagaact 4389
         |||
Qy      4381 aattactaaaacaactcaagacaatctggcagattttaaaaaataacaattcgtt 4440
         |||
Db      4390 aattactaaaacaactcaagacaatctggcagattttaaaaaataacaattcgtt 4449
         |||
Qy      4441 ggaattgtctattatatacaacagattatacaataactattgacaagaattcctagt 4500
         |||
Db      4450 ggaattgtctattatatacaacagattatacaataactattgacaagaattcctagt 4509
         |||
Qy      4501 acagatctgttcttgaanaactctgtcctaaaacgctttatctaaattctactgtacggaac 4560
         |||
Db      4510 acagatctgttcttgaanaactctgtcctaaaacgctttatctaaattctactgtacggaac 4569
         |||
Qy      4561 ttgcaagtgatgttaaacatttcaaacacccaatgcgtlaaaaaacaatgtccacaat 4620
         |||
Db      4570 ttgcaagtgatgttaaacatttcaaacacccaatgcgtlaaaaaacaatgtccacaat 4629
         |||
Qy      4621 tctggaatgttccagacatttagatagaagaagaagaatgtaaatgtttattaaattacaat 4680
         |||
Db      4630 tctggaatgttccagacatttagatagaagaagaagaatgtaaatgtttattaaattacaat 4689
         |||
Qy      4681 caagaagtgatataatgtgttgaanaactccaaatcctactgttaacgagaataatgtgtga 4740
         |||
Db      4690 caagaagtgatataatgtgttgaanaactccaaatcctactgttaacgagaataatgtgtga 4749
         |||
Qy      4741 tctggaatgttccagacatttagatagaagaagaagaatgtaaatgtttattaaattacaat 4800
         |||
Db      4750 tctggaatgttccagacatttagatagaagaagaagaatgtaaatgtttattaaattacaat 4809
         |||
Qy      4801 tctggaatgttccagacatttagatagaagaagaagaatgtaaatgtttattaaattacaat 4860
         |||
Db      4810 tctggaatgttccagacatttagatagaagaagaagaatgtaaatgtttattaaattacaat 4869
         |||
Qy      4861 aactcttaagaatatacctcttaataataactcagtttaataatatacagtttcaat 4920
         |||
Db      4870 aactcttaagaatatacctcttaataataactcagtttaataatatacagtttcaat 4929
         |||

```

RESULT 2
AANS0530
ID AANS0530 standard; DNA: 5760 BP.
XX
XX
AANS0530:

30-SEP-1991 (first entry)
XX
XX
DE Sequence encoding the P195 protein of Plasmodium falciparum.
XX
KM Malaria vaccine: epitope; antigen; immunogen; ss.
OS Plasmodium falciparum.

```

XX      Key      Location/Qualifiers
FH      CDS      216..5179
FT      /*tag= a
PN      EP154454-A.
PD      11-SEP-1985.
PE      21-FEB-1985; 85SEP-0301173.
PR      26-SEP-1984; 84GB-0024340.
PR      22-FEB-1984; 84GB-0004692.
PR      21-FEB-1985; 85GB-0004429.
XX
PA      (WELL ) WELLCOME FOUNDATION LTD.
PI      Holder A, Sandhu J, Odink K, Lockyer M, Riveros-Moreno V;
XX
DR      WPI; 1985-224845/37.
DR      P-PSDB; AAP50777.
XX
PT      Cloned DNA sequence encoding plasmodium falciparum protein -
PT      useful for expressing the protein for use in vaccines against
PT      malaria
XX
PS      Claim 2; Fig 1; 51pp; English.
XX
CC      The sequence encoding the P195 protein of Plasmodium falciparum
CC      (AANS0530) and a peptide comprising at least one of its epitopes
CC      (see AAP50777) are claimed. Also claimed is a vaccine for inducing
CC      immunity to malaria comprising the novel peptide or P195 or a
CC      peptide comprising at least one epitope when derived from the new
CC      DNA sequence, together with a carrier.
XX
SQ      Sequence 5760 BP; 2565 A; 630 C; 725 G; 1840 T; 0 other:

Query Match      97.1%; Score 4779.2; DB 6; Length 5760;
Best Local Similarity 98.4%; Pred No. 0;
Matches 4886; Conservative 0; Mismatches 33; Indels 47; Gaps 4;

Qy      1 atgaagatcatatcttcttattgttcaattctcttttttattataataacacaatgtgta 60
Db      216 atgaagatcatatcttcttattgttcaattctcttttttattataataacacaatgtgta 275
Qy      61 acacatgaaagtatatacagaactgttcaaaaaactagaagctttagaagatgcagattg 120
Db      276 acacatgaaagtatatacagaactgttcaaaaaactagaagctttagaagatgcagattg 335
Qy      121 acagttatagttatttcaaaaagaaagaatgttatataatgaagaacaagtggaca 180
Db      336 acagttatagttatttcaaaaagaaagaatgttatataatgaagaacaagtggaca 395
Qy      181 gctgttacaactagttacacctgtgttcaaaaggttcagttgcttcaggttggc 240
Db      396 gctgttacaactagttacacctgtgttcaaaaggttcagttgcttcaggttggc 455
Qy      241 tcaagtgcttcaagtggtcagttgtctcaggtgtgctcagttgctcaggtcagtt 300
Db      456 tcaagtgcttcaagtggtcagttgtctcaggtgtgctcagttgctcaggtcagtt 515
Qy      301 gcttcagtggttcagttatccaagtcgtaacaatccctcagaataatcaagtgttca 360
Db      516 gcttcagtggttcagttatccaagtcgtaacaatccctcagaataatcaagtgttca 575
Qy      361 gatgtctaaacttaacgtgtttaaaacacagagtaagaataactgtttaaactataaa 420
Db      576 gatgtctaaacttaacgtgtttaaaacacagagtaagaataactgtttaaactataaa 635
Qy      421 gaactcaaatatcctcaactccttgatttaactaatcatatgtttaactgtgtgataat 480
Db      636 gaactcaaatatcctcaactccttgatttaactaatcatatgtttaactgtgtgataat 695

```

481 attcattggttccaataattattgaatgataatgaagaattgaattatattataa 540
|||||
696 attcattggttccaataattattgaatgataatgaagaattgaattatattataa 755
541 ttaactttatttatttatttatttaagaagcaaaattaaatgattgattgtaattat 600
|||||
756 ttaactttatttatttatttatttaagaagcaaaattaaatgattgattgtaattat 815
601 tctcaaatcccttcaactttaaattcgttgcgaatgtaattagacgtacttaaaact 660
816 tctcaaatcccttcaactttaaattcgttgcgaatgtaattagacgtacttaaaact 875
661 ggtgttcggaatagaagaacattagacaattataagaattgaattgaagaagaatt 720
876 ggtgttcggaatagaagaacattagacaattataagaattgaattgaagaagaatt 935
721 tacatttaaaaaaataaaaaacattagaataatataatgaaattatagaagaattag 780
936 tacatttaaaaaaataaaaaacattagaataatataatgaaattatagaagaattag 995
781 aaaaacattgataaaaaataagaatgcgaactaaagaagaagaagaagaagaagaaga 840
996 aaaaacattgataaaaaataagaatgcgaactaaagaagaagaagaagaagaagaaga 1055
841 gtttcaattgattcttcttatttacaataaaacattagaagaagcacataatttaagaag 900
1056 gtttcaattgattcttcttatttacaataaaacattagaagaagcacataatttaagaag 1115
901 gttttagaagaacgtatttacaactttaaanaaaatgaaacacattagaagaatttactgt 960
1116 gttttagaagaacgtatttacaactttaaanaaaatgaaacacattagaagaatttactgt 1175
961 aagataaattgaaattgaagaatcccccacggccaattcttgcgaattacacccaattcttc 1020
1176 aagataaattgaaattgaagaatcccccacggccaattcttgcgaattacacccaattcttc 1235
1021 ctgtgataagaagaagaagaacgcgaggaacggaagaagaagaagaagaagaagaagaatt 1080
1236 ctgtgataagaagaagaagaacgcgaggaacggaagaagaagaagaagaagaagaagaatt 1295
1081 attaaatttaattatgattgatttacttactgacactggaattagaattactataga 1140
1296 attaaatttaattatgattgatttacttactgacactggaattagaattactataga 1355
1141 gaaataaataaataatattgataatgagtgcaaaagttgaaacaaaggaatcaactgaacc 1200
1356 gaaataaataaataatattgataatgagtgcaaaagttgaaacaaaggaatcaactgaacc 1415
1201 aatgataatccaaatggaatttacttacttcttgcataaagaatttaacaatgcttta 1260
1416 aatgataatccaaatggaatttacttacttcttgcataaagaatttaacaatgcttta 1475
1261 aatgataatccaaatggaatttacttacttacttgcataaagaatttaacaatgcttta 1320
1476 aatgataatccaaatggaatttacttacttacttgcataaagaatttaacaatgcttta 1535
1321 aaaaacattatatttatttatttatttatttatttatttatttatttatttatttatttatt 1380
1536 aaaaacattatatttatttatttatttatttatttatttatttatttatttatttatttatt 1595
1381 aaaaatgaaataaataaattgaatttgaataaattcttacaagaacgatttaagtc 1440
1596 aaaaatgaaataaataaattgaatttgaataaattcttacaagaacgatttaagtc 1655
1441 ttaaatgataatacaaaagaatatagaataattacttaattgaataattatgataagaattc 1500
1656 ttaaatgataatacaaaagaatatagaataattacttaattgaataattatgataagaattc 1715
1501 aataataataatgatttaacttaattcgaagaataatgattggttaagaataatcataataa 1560
1716 aataataataatgatttaacttaattcgaagaataatgattggttaagaataatcataataa 1775

1561 gttgagaacttacacacataataacttttgcacccatagaataattcaacataactt 1620
|||||
1776 gttgagaacttacacacataataacttttgcacccatagaataattcaacataactt 1835
1621 gaaaagttaacaaaagcctttaataatataatgagaagatttctttaaagatatagatt 1680
1836 gaaaagttaacaaaagcctttaataatataatgagaagatttctttaaagatatagatt 1895
1681 gaaaagaattaaataatataataaatttaataagaagaataatgagaatgataaga 1740
1896 gaaaagaattaaataatataataaatttaataagaagaataatgagaatgataaga 1955
1741 ttaagtgaataataataaagaatgagaacgcttttgaagaagaagaataacttaagaagc 1800
1956 ttaagtgaataataataaagaatgagaacgcttttgaagaagaagaataacttaagaagc 2015
1801 gaaaataaaccaagattgaataaattttagaagttcgtgaattgtaaaatgataagttcaa 1860
2016 gaaaataaaccaagattgaataaattttagaagttcgtgaattgtaaaatgataagttcaa 2075
1861 aaggtttatttaattgaacaaatgacgaattaaagaacccaattgattttaaaat 1920
2076 aaggtttatttaattgaacaaatgacgaattaaagaacccaattgattttaaaat 2135
1921 gtagaattaaacataataataatgacgaatttcccaattcttacaagaagaataagcaaga 1980
2136 gtagaattaaacataataataatgacgaatttcccaattcttacaagaagaataagcaaga 2195
1981 ccttattatttaattggttgaagaagaagaattgataaattaaagtgttcacgtcgaag 2040
2196 ccttattatttaattggttgaagaagaagaattgataaattaaagtgttcacgtcgaag 2255
2041 gtagaattcatttgaataatgagaagaagaagaataaataacgaagaatgcaatcggaatt 2100
2256 gtagaattcatttgaataatgagaagaagaagaataaataacgaagaatgcaatcggaatt 2315
2101 tcggaacccatcaacgcgaagaagaagaataaagaagcaagcaacttacaacacgtgacacaa 2160
2316 tcggaacccatcaacgcgaagaagaagaataaagaagcaagcaacttacaacacgtgacacaa 2375
2161 gcaggaattcgttgaagaagaagaattcagtaacgaagcaagcaagcaagcaagcaagca 2220
2376 gcaggaattcgttgaagaagaagaattcagtaacgaagcaagcaagcaagcaagcaagca 2435
2221 caaccacccagttaccagttaccagttaccagttaccagttaccagttaccagttaccag 2280
2436 caaccacccagttaccagttaccagttaccagttaccagttaccagttaccagttaccag 2495
2281 ccagtaataataaataaacttgaataatggttcccaattgaattatcttgaagaataataga 2340
2496 ccagtaataataaataaacttgaataatggttcccaattgaattatcttgaagaataataga 2555
2341 tttttaaatcttcataatataatgltccaaatatatttgggttccaaactcaactgaac 2400
2556 tttttaaatcttcataatataatgltccaaatatatttgggttccaaactcaactgaac 2615
2401 gaaaagaatttaaaacataataaatttacaagaaggaagaagaagaagaagaagaatt 2460
2616 gaaaagaatttaaaacataataaatttacaagaaggaagaagaagaagaagaagaatt 2675
2461 gatccattgactatttgggttaataatacaaaataacatcccttaattgattcattgatt 2520
2676 gatccattgactatttgggttaataatacaaaataacatcccttaattgattcattgatt 2735
2521 gatagcttaaacatagatttaccacactat----- 2551
2736 gatagcttaaac- atagtttaccacactattatggtttatgaaagaagaattggtta 2794
2552 -----ttagggaattttagaagaagaagaattggtttgatttataataact 2597
2795 tgtatttataattcattatggaataattttagaagaagaagaattggttttataataact 2854
2598 taaggaataatgacaaaatttaaaatttatttagggaagcgaaaaagttatccacattgt 2657

Db 2855 taaggataagcaaaatttaaaattatttaggaagcgaaaaaglatccacatcgt 2914
OY 2658 aaaaactcttcaagttacatcgaatgcacattatcatatcaacccagataaccga 2717
Db 2915 aaaaactcttcaagttacatcgaatgcacattatcatatcaacccagataaccga 2974
OY 2718 agtaagtcgaactgagatcatcacaatctcacaatttgtaaatagtttaaatatc 2777
Db 2975 agtaagtcgaactgagatcatcacaatctcacaatttgtaaatagtttaaatatc 3034
OY 2778 tgaacaacatcgaatccttgcgaaaaaaataataaccgaagttaagtcgaaa 2837
Db 3035 tgaacaacatcgaatccttgcgaaaaaaataataaccgaagttaagtcgaaa 3094
OY 2838 aagtagtgaanaactttatgaanaagatttaanaagtagtgcacatttatgatc 2897
Db 3095 aagtagtgaanaactttatgaanaagatttaanaagtagtgcacatttatgatc 3154
OY 2898 tttaacaattttgttaaaatctaaagctgataatattatctatgcattgcacaa 2957
Db 3155 tttaacaattttgttaaaatctaaagctgataatattatctatgcattgcacaa 3214
OY 2958 aaggaagaaattgaagaagataatataaataaaaaaacttcaagttacattga 3017
Db 3215 aaggaagaaattgaagaagataatataaataaaaaaacttcaagttacattga 3274
OY 3018 ttatataataatataaataatataaattagaagaattttgttaaaaaagaacgttgg 3077
Db 3275 ttatataataatataaataatataaattagaagaattttgttaaaaaagaacgttgg 3334
OY 3078 taataataaatgcgaatttaaaaaacttactttataaagaacaaattgaacaaact 3137
Db 3335 taataataaatgcgaatttaaaaaacttactttataaagaacaaattgaacaaact 3394
OY 3138 gaattcaacttaatacccaagcatglatcaaaaacttctgttcttcttaacaaa 3197
Db 3395 gaattcaacttaatacccaagcatglatcaaaaacttctgttcttcttaacaaa 3454
OY 3198 aaaaagagcttgaatacgcaagaacttgaanaacacattagaanaaacaataatattga 3257
Db 3455 aaaaagagcttgaatacgcaagaacttgaanaacacattagaanaaacaataatattga 3314
OY 3258 acatttaagaagacttgaataatataatggtgtaatccctccttaaaaactttaag 3317
Db 3515 acatttaagaagacttgaataatataatggtgtaatccctccttaaaaactttaag 3574
OY 3318 tgaagaatcaattcaaacgaagaataatattgcagtttagaanaactttaaglattag 3377
Db 3575 tgaagaatcaattcaaacgaagaagaataatattgcagtttagaanaactttaaglattag 3634
OY 3378 taatttagaagaanaatttaagaagataatttaattagaanaagaanaaattatcactt 3437
Db 3635 taatttagaagaanaatttaagaagataatttaattagaanaagaanaaattatcactt 3694
OY 3438 atcaagtggaattacatcatttaattgtcgaatttaagaagaagtataaaaaataaattta 3497
Db 3695 atcaagtggaattacatcatttaattgtcgaatttaagaagaagtataaaaaataaattta 3754
OY 3498 taacaggttaattcccaagtgaanaataatacgaagttaacaatgcattgaacttataca 3557
Db 3755 taacaggttaattcccaagtgaanaataatacgaagttaacaatgcattgaacttataca 3814
OY 3558 aaaaattctcccaagaagaagatggttcgaacagttgttaagtgaagtgcgcgaac 3617
Db 3815 aaaaattctcccaagaagaagatggttcgaacagttgttaagtgaagtgcgcgaac 3874
OY 3618 attagaacaagtcgaacgaagaacacagacatcactgttagagcgaggtctaacac 3677
Db 3875 attagaacaagtcgaacgaagaacacagacatcactgttagagcgaggtctaacac 3934
OY 3678 aataacaacatcacaaatgtcgaatgaagaatgaagaatgaagaatgaagaatgaagaat 3737
Db 3935 aataacaacatcacaaatgtcgaatgaagaatgaagaatgaagaatgaagaatgaagaat 3994
OY 3738 tggagaatccgaagaagattatgatatttaggaagaagtgaacgggaagaacgaataac 3797
Db 3995 tggagaatccgaagaagattatgatatttaggaagaagtgaacgggaagaacgaataac 4054
OY 3798 tcccttcgtaatttgataacatacttcttaaaatttgaaatgtaataagtttataattt 3857
Db 4055 tacttcgtaatttgataacatacttcttaaaatttgaaatgtaataagtttataattt 4114
OY 3858 aaaaacttgaacagtggttataagaagtttaaaaaaacaattgaanaataagtttgcac 3917
Db 4115 aaaaacttgaacagtggttataagaagtttaaaaaaacaattgaanaataagtttgcac 4174
OY 3918 attaatgttaagtgttaagaagatttaattcacagatttaaaacggtgaanaatttcaa 3977
Db 4175 attaatgttaagtgttaagaagatttaattcacagatttaaaacggtgaanaatttcaa 4234
OY 3978 aatgtttagaatacgaatttaattccataaagaatttaacaataagtaattgtgt 4037
Db 4235 aatgtttagaatacgaatttaattccataaagaatttaacaataagtaattgtgt 4294
OY 4038 caaagatccataaatttcttaataaagaaaaaagagataaattcttaagcagttataa 4097
Db 4295 caaagatccataaatttcttaataaagaaaaaagagataaattcttaagcagttataa 4354
OY 4098 ttatataagatttcaatgatacagataaattttgcgaatgagttcttgcgataat 4157
Db 4355 ttatataagatttcaatgatacagataaattttgcgaatgagttcttgcgataat 4414
OY 4158 taataatataccgaanaaataataatcagaatttagattcaatttaaaaaataatacga 4217
Db 4415 taataatataccgaanaaataataatcagaatttagattcaatttaaaaaataatacga 4474
OY 4218 caaacaagtggaatggaataacacttcccttttaaacaataatgagacattataa 4277
Db 4475 caaacaagtggaatggaataacacttcccttttaaacaataatgagacattataa 4534
OY 4278 aacagttatgataaatttgattttgttaatttcttaagaagcaaaagtttaataa 4337
Db 4535 aacagttatgataaatttgattttgttaatttcttaagaagcaaaagtttaataa 4594
OY 4338 taacatlggaataatcaacagtagaagtttaaaataaagaacattatcttaaaaaaat 4397
Db 4595 taacatlggaataatcaacagtagaagtttaaaataaagaacattatcttaaaaaaat 4654
OY 4398 tcaagaacaattgycagattttaaaaaataaacaatttcogtttggaattgctgattatc 4457
Db 4655 tcaagaacaattgycagattttaaaaaataaacaatttcogtttggaattgctgattatc 4714
OY 4458 aacagttataacataaattatcttgcgaagaagttcccttagaagcagatgatttga 4517
Db 4715 aacagttataacataaattatcttgcgaagaagttcccttagaagcagatgatttga 4774
OY 4518 aactctgtcaaaacggtttatcttaatttacttga-tgnaaactgcgaaggttga 4576
Db 4775 aactctgtcaaaacggtttatcttaatttacttga-tgnaaactgcgaaggttga 4834
OY 4577 acatttcaacaacccaatgcgttaaaaaaacaatg--tccacaacaattctgagattcag 4634
Db 4835 acatttcaacaacccaatgcgttaaaaaaacaatgccaacagagttctcgagattcag 4894
OY 4635 acatttagatgaagaagaagatgtaattgttataatataatacaacaagaagtgataa 4694
Db 4895 acatttagatgaagaagaagatgtaattgttataatataatacaacaagaagtgataa 4954
OY 4695 atgtgttgaanaatcccaacttcaactgtgaacgaanaataatggttgcgtgcgaatgc 4754
Db 4955 atgtgttgaanaatcccaacttcaactgtgaacgaanaataatggttgcgtgcgaatgc 5014
OY 4755 caaatgtaccgaagaagattcagtgcaacggaagaagaataatcacaatgtaattactaa 4814
Db 5015 caaatgtaccgaagaagattcagtgcaacggaagaagaataatcacaatgtaattactaa 5074

QY 4815 accgatcttaccacatttcgagtattcttcgcagttctcctaacttcttagaat 4874
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 5075 accgatgtttaccacatttcgagtattcttcgcagttctcctaacttcttagaat 5134
QY 4875 atcatcttataactacatgcttaataataatacagttccatttaa 4920
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 5135 atcatcttataactacatgcttaataataatacagttccatttaa 5180

RESULT 3

AAV21451

ID AAV21451 standard; DNA; 4940 BP.

XX AC AAV21451;

XX DT 23-SEP-1998 (first entry)

XX DE P. falciparum modified gp190 DNA.

XX gp190: malaria: MSP-1; merozoite surface protein; stability; vaccine;
KM monoclonal antibody; passive immunisation; parasite: ss.

XX Plasmodium falciparum.
OS Synthetic.

XX Key Location/Qualifiers
FH CDS 10..4928
FT /*tag= a
FT /product= gp190

XX PN MO9814583-A2.

XX PD 09-APR-1998.

XX PF 02-OCT-1997; 97MO-EP05441.

XX PR 02-OCT-1996; 96DE-4040817.

XX PA (BUJA/) BUJARD H.

XX PI Bujard H, Pan W, Tolle R;

XX DR MPI; 1998-240088/21.

XX DR P-PSDB; AAM54145.

XX PT Recombinant production of complete gp190/MSP-1 Plasmodium surface
XX protein - useful in anti-malaria vaccines, also stabilising genes by
XX reducing their AT content

XX PS Example 1; Fig 3c; 48pp; German.

XX This sequence encodes a modified Plasmodium falciparum gp190/MSP-1
XX (merozoite surface protein) which has a reduced AT content resulting in
XX a higher stability of the protein. Such a protein is useful in
XX vaccines against malaria or for producing monoclonal antibodies (for
XX passive immunisation). The complete gp190 protein can now be produced
XX outside the parasite and has, at least over extended regions, the native
XX pattern of folding. Larger amounts of the protein can be produced
XX recombinantly than would be possible using the parasites as source.

XX Sequence 4940 BP; 1669 A; 1156 C; 1060 G; 1055 T; 0 other;

Query Match 57.2%; Score 2812.8; DB 19; Length 4940;

Best Local Similarity 73.2%; Pred. No. 1.3e-312;

Matches 3603; Conservative 0; Mismatches 1317; Indels 0; Gaps 0;

QY 1 atgaagatcatattctttatgctcattctttttttatataataacacaaagtga 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 10 atgaataatattctctctctcttcttcttcttcttcttcttcttcttcttctg 69
QY 61 acacatgaagtatcaagaactgtcacaataaactagaagctttagaagatgcagatg 120

Db 70 accacagaatccatcatcagaagcgtgttaagaacacttgaagcttggaaagtccgcctt 129
QY 121 acaggttatagttatttcaaaaggaanaaatgttataatgaagaaacagttgaaca 180
Db 130 accggaataagcgtgttccagaaggaagatgtgtctgaatgaaggaagcagttgcacg 189
QY 181 gctgttaacactagtaaccctgttcaaggttcaaggttgcagttgcaggtgtgttcaggtgc 240
Db 190 gccgttacacaaccgacaccgcgttctcaaaaggtctgtgcagcgtgtgtcccggtgg 249
QY 241 tcaagttgtcaggtgtgcagttgttcaggtgtgtcaggtgtgtcaggtgtgtcaggt 300
Db 250 tctgtgcctcttggggttccgttcgttcgttcgttcgttcgttcgttcgttcgttcgttc 309
QY 301 gcttcaggtgtgtcaggttcaagaaatcgaacgttacaaccccttcagtaataatcaagtttca 360
Db 310 gcaagcgcgtgtccggaacagtcgaagaacccaatccatctgcacaactcagcagttcc 369
QY 361 gatgttaactttagcgttatttaaaacacagagttacgaattactgtttaactacaaa 420
Db 370 gacgccaaagctcctacgcgcgaacctcaagcacgcgagtgaagaactatctcctcactacaag 429
QY 421 gaactcaaatctcctaactcttgaatttaactaactatgttaacttgtgtgataat 480
Db 430 gagctgaagtaaccacaggtgttgcacctcaactaactatcatgtgcagctgtgtgatac 489
QY 481 attcatgttccaataatttaattgtatgatatgaagaatgaagaatgaagaatgaagaat 540
Db 490 attcatgttccaataatttaattgtatgatatgaagaatgaagaatgaagaatgaagaat 549
QY 541 ttaaaactttatttgatttatttaagaagaanaatttaagaattgaattgaattgaattgaatt 600
Db 550 ttgaattcttactctgacttgcaggaagccaacttgaatgaatgttgcgcgaactgactac 609
QY 601 tgcacaataccttcaactcttaaatctgtcgaatgaatgaatgaatgaatgaatgaatgaat 660
Db 610 tgcacaataccttcaactcttgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 669
QY 661 gttctcgatatagaanaaccatagacaatatttaagaatgttagaanaatgtgaagat 720
Db 670 gttctcgatatagaanaaccatagacaatatttaagaatgttagaanaatgtgaagat 729
QY 721 tacatttaaaaaaataaaaaaacctagaaaataataagaatttaagaagaatgaag 780
Db 730 tacatttaaaaaaataaaaaaacctagaaaataataagaatttaagaagaatgaag 789
QY 781 aaacacattgataaataaataagaatgcaactaagaagaagaanaaaaaaataatatac 840
Db 790 aagacacatagacaanaataagaatgcaactaagaagaagaagaagaagaagaatgtgtac 849
QY 841 gctcaatatgattcttctatttaacaataaacttagaagaagcacataatthaataagc 900
Db 850 gccacgtacgacctgtccactataacaacagcttgaagaagcccatcactacacagc 909
QY 901 gttttgaanaaagttatgacactttaaanaaaaaatgaanaacatttaagaattactgtg 960
Db 910 gtacttgaagaagcgtcagtagacacccctcagaagaatgaagaatatacaagaactgcgcagc 969
QY 961 aagataatgaatlaaataatcccccacgcgcgaacttgcgaatcacacaatacttc 1020
Db 970 aagatlaatgaatlaaataatcccccacgcgcgaacttgcgaatcacacaatacttc 1029
QY 1021 cttgataagaacaaaataatcgaagacacgaanaaataaagaatctgcacaact 1080
Db 1030 cttgataagaacaaaataatcgaagacacgaanaaataaagaatctgcacaact 1089
QY 1081 attaaatttaactatgatatgttattactatgacacttgaatgaataatcattataaga 1140
Db 1090 attaaatttaactatgatatgttattactatgacacttgaatgaataatcattataaga 1149
QY 1141 gaaaaaataaataatgtatataagttgcaaggttgaagaagaagaatcaactgaacc 1200

Dh 1150 gagaaagataagataatgacatctccgcaaaagtcgagacaagaatcaacggaacct 1209
Qy 1201 aatgaatataccaataggaattactctcttgctcatatagaatattaaacgtctta 1260
Dh 1210 aatgaatataccaataggaattactctcttgctcatatagaatattaaacgctctc 1269
Qy 1261 aatgaatataccaataggaattactctcttgctcatatagaatattaaacggaacct 1320
Dh 1270 aatgaatataccaataggaattactctcttgctcatatagaatattaaacggaacct 1329
Qy 1321 aaaaacataataatgataaagaagaaaaatctataaataaataaataaataaataa 1380
Dh 1330 aagaataataataaataaataaataaataaataaataaataaataaataaataaataa 1389
Qy 1381 aaaaataaataaataaataaataaataaataaataaataaataaataaataaataa 1440
Dh 1390 aaaaataaataaataaataaataaataaataaataaataaataaataaataaataa 1449
Qy 1441 ttaaatatataaataaataaataaataaataaataaataaataaataaataaataaataa 1500
Dh 1450 ctataagataataataaataaataaataaataaataaataaataaataaataaataaataa 1509
Qy 1501 aataataataataaataaataaataaataaataaataaataaataaataaataaataa 1560
Dh 1510 aataataataataaataaataaataaataaataaataaataaataaataaataaataa 1569
Qy 1561 gtggaataaataaataaataaataaataaataaataaataaataaataaataaataa 1620
Dh 1570 gtggaataaataaataaataaataaataaataaataaataaataaataaataaataa 1629
Qy 1621 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 1680
Dh 1630 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 1689
Qy 1681 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 1740
Dh 1690 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 1749
Qy 1741 ttaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 1800
Dh 1750 ctggtatataaataaataaataaataaataaataaataaataaataaataaataaataa 1809
Qy 1801 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 1860
Dh 1810 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 1869
Qy 1861 aaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 1920
Dh 1870 aaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 1929
Qy 1921 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 1980
Dh 1930 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 1989
Qy 1981 aaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2040
Dh 1990 aaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2049
Qy 2041 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2100
Dh 2050 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2109
Qy 2101 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2160
Dh 2110 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2169
Qy 2161 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2220
Dh 2170 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2229
Qy 2221 aaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2280
Dh 2230 aaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2289

Qy 2281 caaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2340
Dh 2290 caaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2349
Qy 2341 ttaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2400
Dh 2350 ttaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2409
Qy 2401 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2460
Dh 2410 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2469
Qy 2461 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2520
Dh 2470 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2529
Qy 2521 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2580
Dh 2530 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2589
Qy 2581 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2640
Dh 2590 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2649
Qy 2641 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2700
Dh 2650 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2709
Qy 2701 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2760
Dh 2710 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2769
Qy 2761 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2820
Dh 2770 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2829
Qy 2821 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2880
Dh 2830 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2889
Qy 2881 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2940
Dh 2890 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 2949
Qy 2941 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 3000
Dh 2950 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 3009
Qy 3001 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 3060
Dh 3010 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 3069
Qy 3061 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 3120
Dh 3070 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 3129
Qy 3121 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 3180
Dh 3130 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 3189
Qy 3181 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 3240
Dh 3190 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 3249
Qy 3241 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 3300
Dh 3250 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 3309
Qy 3301 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 3360
Dh 3310 gaaaatataaataaataaataaataaataaataaataaataaataaataaataaataa 3369

QY 1882 attgacgnaattaaaaagactcaattgattttaaaaaatgtagaattaaaaataata 1941
 DB 1978 atagaagacttaagaagaattgattattttttaaaaaatgcacactaaaaagtagatt 2037
 QY 1942 catgttcccaattcttaacaaacagaataaagaagaaacctattatttaattgtgt 2001
 DB 2038 catgttcccaattcttaacaaacagaataaagaagaaacctattatttaattgtgt 2097
 QY 2002 aaaaaagaattgataaatttaaaagtgtcatgtcttaagttagaattatgataatgaa 2061
 DB 2098 aaaaaagaattgataaatttaaaagaatttataccaaaagtaaaagacatgttaaaagaa 2157
 QY 2062 gaaaaaaaaaacataaaacagaagtgcaatcgataatttcggaacctcaacccagaagaa 2121
 DB 2158 gaacaaagctgtctatcaatgattatacaaacctttagttgtcagcaagcgaaac---aact 2214
 QY 2122 gaaatacagagcaagcaactacaaaacctgtgacaacaaagcagatctgtcttagaagaa 2181
 DB 2215 gaagatggggtgtcaatccacacacacattatcccaatcagagaagaagaaagtaacagaa 2274
 QY 2182 gattcgttcaacgacacagacacagaacaaaacaaagcacaacccacagttacagttacca 2241
 DB 2275 gaaacagaagaacacagaaagaaacagtaggacacacaaacaggttaacaaataacattacca 2334
 QY 2242 gtacccagaagcaaaagcacacagtcaccaacacagcagcagcagtaataataaactgtgaa 2301
 DB 2335 ccaaaagaagtaaaagtgtttgaaatttcaatagaacataagaatgataatgacaattacaa 2394
 QY 2302 aatgtttccaattgatacttcttgaanaattatagtaattttaaatactatataata 2361
 DB 2395 gcttgcataaaaacagttatcttaagaagaattagatgaatttttaactaaatcatataata 2454
 QY 2362 tgcacacaaatattttgttttcaacactcaactgaagcagaagaagaatataaacaatat 2421
 DB 2455 tgcataaataatattttgattcaaaactctagatagtcacaaaataattatagaggtatata 2514
 QY 2422 aaattacaaaaggaagaaagaacaaatgaattgatactgtatcattatagactattgttt 2481
 DB 2515 aattcttaccocgaagaagaagaataatgaataaataatcatgtatcccatagattattatttt 2574
 QY 2482 aatatacaaaatatacactgtatgtatcttctatgttttgaatgacttaaaactaagttta 2541
 DB 2575 aatttcaaaatatacactgtatgtatcttctatgtatgattatgataagtaagcaatgtttta 2634
 QY 2542 tcccaactattttatgaaattttatgaaaaagaatgtgttgaattatataaacttaag 2601
 DB 2635 caacactctcttttgaattatatacaaaaggaatgtattattatatacctaaactaaa 2694
 QY 2602 gataatgacaaaatttaaaaatttatagagaagcgaaaaa-----gtatcacactct 2655
 DB 2695 gaaagaataacacatcaaaaataattatagagaagcagaacaaaataaactgtgaacatcatct 2754
 QY 2656 gtaaaaaacttctcaagtctatcaatgcaacattatcatlaacactcaggaataaacc 2715
 DB 2755 acatccagctcctggaataatacaccgtaaatatgtctcaatccgcaactcagtaattcc 2814
 QY 2716 ga----- 2717
 DB 2815 caaaaccaaactcaaatgcatctccttaccataaccnaaatgtgttagctgtatcatct 2874
 QY 2718 -----agtaagtgcataatgatacattacatcttcaaaattgtaataagt 2766
 DB 2875 ggtcccggtgtgttgaaagaagtcataatcccttaacagttatgtgtatagtaacgat 2934
 QY 2767 ttaaaattttgaaacaaatgtgagcttctgtaaaaaacaaataatatacagaagattta 2826
 DB 2935 ttgaagaagtgtgttagtctcttaaatcttggaaataaaactaaagtaactaatcatcat 2994
 QY 2827 ----ataggtcacaagaagtgtaaaactttatagaagaatataaagaatagttatata 2883
 DB 2995 accatttctacacagagatgtaaaaaattttatgagaatatttttaaaaaataatgatacc 3054
 QY 2884 tttaataatgaatcttttaacaatttctgaanaattcttaagaagctgataatataattcatgt 2943

DB 3055 tattttaatgatatcaaacacttcgttaaaacttcaattccaagaagtaattacaggtttg 3114
 QY 2944 aatgtatgatacaaaaaggaagaattagaagaagatatttaataatataaaaaacttta 3003
 DB 3115 a---ccgaacacacaaaaatgcatlaaatgtatgaatttataaaaaattaaagaatacttta 3171
 QY 3004 cagttatcatgtattatataataataataataataaattagaagaatatttggtaaa 3063
 DB 3172 cagttatcatgtattatataataataataataataaattagaatgattatttaataag 3231
 QY 3064 aagaacacagttgttataataataaalgcaaatlaaaaaacttcttattataaagaacaa 3123
 DB 3232 aaaaaagaacttggccaagcacaatgtcaaatlaaaaaacttcttattataaagaacaa 3291
 QY 3124 ttgaatccaatattgatacttaacttaatacccaagcatgtattacaaaactttctgt 3183
 DB 3292 ttgaatccaatattgatacttaacttaatacccaatgtattataaaaactttctgt 3351
 QY 3184 ttctttaacaaaaaagaagctgtaaatatagcagaactgtgaacacattagaacaca 3243
 DB 3352 ttctttaacaaaaaagaagctgtaaatatagcagaactgtgaacacattagaacaca 3411
 QY 3244 aaaaattatgtaaacatttataagaagctgttataattatataatgtgtatcatctcca 3303
 DB 3412 aaaaattatgtaaacatttataagaagctgttataattatataatgtgtatcatctcca 3471
 QY 3304 ttaaaaacttgaatgaagaatccaactccaacagaagaataattatgtccagtttagaanaac 3363
 DB 3472 ttaaaaacttgaatgaagaatccaactccaacagaagaataattatgtccagtttagaanaac 3531
 QY 3364 tttaagaattatgaatgaagaagaataatagaagaatatttaattatagaagaaga 3423
 DB 3532 tttagatgataagtaaaatagaatgaatgaagaatgaagaatatttcaactttagaagaaga 3591
 QY 3424 aaattatcaacttcaagtgatcatcatcatcttattgtctgaatttaaaagaagtaata 3483
 DB 3592 aaattatcttcttcaacagtgatcatcatcatcttattgtctgaattttaaagaagtaata 3551
 QY 3484 aaaaataaaaattatacaggttaattctccaagtgaanaatacagatgttaacaatgca 3543
 DB 3552 aaaaataaaaattatacaggttaattctccaagtgaanaatacagatgttaacaatgca 3711
 QY 3544 ttgaatcttcaaaaaaattcttcccaagaagaagcagat----- 3582
 DB 3712 ttaaaatcttcaagaaatttcttcccaagaagcagaagtttaacaagttgtaactcaact 3771
 QY 3583 -----gttgcacagttgtgaagtgaagtgatccgacacttagaa 3624
 DB 3772 caaccgagtgttaactccatctcattatctgttaaggtgaagtgtgttcagatccaca 3831
 QY 3625 caaagtcaaccgaagaagaacagcatcaactcatgttaggcagagttcaacacataaaca 3684
 DB 3832 aagaagaagaacacaaatacactcagctcttatttaacagaattatacaacaagtagta 3891
 QY 3685 acatcaacaaatgtcagatgagtagatgaatgaatgaatgaatgaatgaatgaatgaatga 3744
 DB 3892 caattcaaaaattatagcagaagaatgtatcttgaattgaattacccatttggagaa 3951
 QY 3745 tccgaagaagattatgatttaggaagaagtagtaacaggaagaagcagtaactctcc 3804
 DB 3952 tccgaagaagattatgagaaatatttagatcaagtagtaactcggagaagcagaa---ctgtgc 4008
 QY 3805 gtaattgataacatcttcttaaaattgaaatgaatgaatgaatgaatgaatgaatgaatga 3864
 DB 4009 acatgtgataatctcctccaggtattgaaatgaatgaatgaatgaatgaatgaatgaatga 4068
 QY 3865 tttagcaggtgttataagaagtttaaaaaacaaattagaanaataaagtttgcacattta 3924
 DB 4069 tttagcaggtgttataagaagtttaaaaaacaaattagaanaataaagtttgcacattta 4128
 QY 3925 gttaaagtttaagatattttaaattcaagatttaaaacgtgaanaattcaaaaatgtt 3984

```

Db 4129 ttaattgacgatalcatltaaatlcacgcttccttaagaacacaaatatctttagatgta 4188
QY 3985 ttagaatcagatttaattccatcatalaagatttaacatcagtaattatgtgtcacaagat 4044
Db 4189 ttagaatcagatttaagcaatttaaacatataatcccaaatgatacatatgtaagat 4248
QY 4045 ccatataaattccttaaaagaaaagagataaattccttaagcaatttaataatc 4104
Db 4249 tcatataattatgtaattcagaacaaaacacacttttaaaagttaacaatatata 4308
QY 4105 aagatcacaatagatcagatataaatttgcagaatgtcttcttgatattataata 4164
Db 4309 aagaaatcagtaaaatgtaatttaatttgcacaggaagtataagttatgaaag 4368
QY 4165 ttatccgaaaatataatcagattttagatcatttaaaaaatatcaaa----- 4214
Db 4369 gtttagcgaatatataagatgatttagaacatataaaaagttaacaaagaagaag 4428
QY 4215 -----cgacaacaa 4224
Db 4429 gagaagttcccatatcaccacacaacacctccgcacagtaaaacagagaacaa 4488
QY 4225 gctgaaaatgagaatcccttcccttttaacaataatgagacctatataaacagtt 4284
Db 4489 aagaaggaagtagtcccttccattttaacaaacatlgagacctatataacattta 4548
QY 4285 aagataaattgatttatttgaattcatttagaagcaaaagtctcaaatatatacat 4344
Db 4549 gttataaatttgcagttcttcttaatttaactaaaggcaaaagattacagattgta 4608
QY 4345 gagaatcaaacgtagaagttaaataaagaacttaattacttaaaaaaatcaaacag 4404
Db 4609 gaataaagatgaagacatgtaataataacttaacttagattgatttaaaagaattgag 4668
QY 4405 aaattgcaagattttaaaaaaataaacaatttgcgttgagattctgattatacagat 4464
Db 4669 aaaaatagatttttaaaaacccataacgacttgcgaagcaatttaaaaaattgataa 4728
QY 4465 tataaccataataactattgacaaagtctccttagtaacagatgatttgaanaactt 4524
Db 4729 gatacgaaaaaagatagctgtgcgaattacttgaacagagattag---ttcaaaattt 4785
QY 4525 gctaaaaacgcttattcatttacttactgtgaaacttgcgaaggtatgttaaacattca 4584
Db 4786 cctaatacaataatcaaaaatattgaaggaatattccaagatatgttaaacattca 4845
QY 4585 caacaccaatgctgtaaaaaaacaatgtccacaacaaattctgagattcagacattgat 4644
Db 4846 caacaccaatgctgtaaaaaaacaatgtccagaanaattctgagattcagacattgat 4905
QY 4645 gaagaagaagaatgtaatttatttaataattacaacagaagaagtgttaagtgtgaa 4704
Db 4906 gaagaagaagaatgtaatttatttaataattacaacagaagaagtgttaagtgtgaa 4965
QY 4705 aatccaatctacttgcgaagaataatggtgagatgtaacagatgcgaatgtacc 4764
Db 4966 aatccaatctacttgcgaagaataatggtgagatgtaacagatgcgaatgtacc 5025
QY 4765 gaagaagatcaggtagcaacggaagaanaatcacatgtgagttactaaacctgattc 4824
Db 5026 gaagaagatcaggtagcaacggaagaanaatcacatgtgagttactaaacctgattc 5085
QY 4825 tatcactttgagtgattttctgcagttcctcttaacttcttagaataatatttta 4884
Db 5086 tatcactttgagtgattttctgcagttcctcttaacttcttagaataatatttta 5145
QY 4885 ttaatacagtttaataatataacagttcattta 4920
Db 5146 ttaatacagtttaataatataacagttcattta 5181

```

RESULT 5
AAT97956

```

ID AAT97956 standard; DNA; 1950 BP.
XX
AC AAT97956;
XX
DT 03-APR-1998 (first entry)
XX
DE Chimeric MSA-1 antigenic protein 1 used in a malaria vaccine.
XX
KW Signal peptide; malaria; vaccine; merozoite surface antigen-1 peptide;
KW MSA-1; recombinant vaccinia virus; Plasmodium falciparum; anchor peptide;
KW immune response; humoral; cell-mediated; merozoite; ss.
XX
OS Chimeric - Mammalia.
OS Chimeric - Plasmodium falciparum.
OS Chimeric - Homo sapiens.
XX
FH Key
FH misc_feature
FT 1..165
FT /tag= a
FT /note= "signal sequence added to improve
FT immunogenicity"
FT misc_feature
FT 166..1893
FT /tag= b
FT /note= "Plasmodium falciparum MSA-1 peptide"
FT misc_feature
FT 1894..1950
FT /tag= c
FT /note= "anchor sequence added to improve
FT immunogenicity"
PN WO9726911-A1.
PD 31-JUL-1997.
XX
PF 29-JAN-1997; 97WO-US01395.
XX
PR 29-JAN-1996; 9605-0593006.
XX
PA (GEOU ) UNIV GEORGETOWN.
XX
PI Davidson EA, Yang S;
XX
DR MPI; 1997-393372/36.
XX
PT Malaria vaccine - comprises expression vector expressing fragment of
PT merozoite surface antigen
XX
PS Claim 33; Fig 2; 75pp; English.
XX
CC The present sequence encodes a chimeric protein that contains a
CC mammalian signal and anchor sequence, and a merozoite surface
CC antigen-1 (MSA-1) carboxy terminal peptide. The C-terminal fragment of
CC MSA-1 provides a more specific response than the complete MSA-1,
CC and attachment of anchor and signal sequences improve the immunogenicity
CC of the protein better than the use of an adjuvant. The chimeric protein,
CC and expression vectors (analogues that express MSA-1 without either
CC signal or anchor peptides), particularly in the form of recombinant
CC vaccinia virus, are used in vaccines to prevent or treat malaria caused
CC by Plasmodium falciparum. The vaccinia vector expresses the antigen
CC fragment for many days, or even years, generating a long-lasting immune
CC response (humoral and/or cell-mediated) against the merozoite form of
CC the parasite, in humans or other animals.
XX
SQ Sequence 1950 BP; 830 A; 236 C; 277 G; 607 T; 0 other;

```

Query Match 36.6%; Score 1798.8; DB 18; Length 1950;
Best local Similarity 95.5%; Pred. No. 3.1e-197;
Matches 1862; Conservative 0; Mismatches 87; Indels 1; Gaps 1;

```

QY 2972 aagaagatatataaataaanaaaacttacaagtatcatcattgattatataatnaat 3031
Db 1 atgaagatcatattctttttagtgcattcttctttttattatataatacaacagtgtga 60

```



```

FH Key Location/Qualifiers
FT misc_feature 1..1839
FT /tag= a
FT /note= "Plasmodium falciparum MSA-1 peptide"
FT misc_feature 1840..1897
FT /tag= b
FT /note= "anchor sequence added to improve immunogenicity"
XX
XX MO9726911-A1.
XX
XX 31-JUL-1997.
XX
XX 29-JAN-1997; 97WO-US01395.
XX
XX 29-JAN-1996; 96US-0593006.
XX
XX (GEOU ) UNIV GEORGETOWN.
XX
XX Davidson EA, Yang S;
XX
XX WPI; 1997-393372/36.
XX
XX Malaria vaccine - comprises expression vector expressing fragment of
XX merozoite surface antigen
XX
XX Claim 33; Fig 4; 75pp: English.
XX
XX The present sequence encodes a chimeric protein that contains a
XX human anchor sequence, and a merozoite surface antigen-1 (MSA-1) carboxy
XX terminal peptide. The C-terminal fragment of MSA-1 provides a more
XX specific response than the complete MSA-1, and attachment of the
XX anchor sequence improve the immunogenicity of the protein better than
XX the use of an adjuvant. The chimeric protein, and expression vectors
XX (analogues that express MSA-1 without either signal or anchor peptides),
XX particularly in the form of recombinant vaccinia virus, are used in
XX vaccines to prevent or treat malaria caused by Plasmodium falciparum. The
XX vaccinia vector expresses the antigen fragment for many days, or even
XX years, generating a long-lasting immune response (humoral and/or
XX cell-mediated) against the merozoite form of the parasite, in humans or
XX other animals.
XX
XX Sequence 1897 BP; 815 A; 230 C; 274 G; 578 T; 0 other;
XX
Query Match 36.3%; Score 1785.4; DB 18; Length 1897;
Best Local Similarity 97.2%; Pred. No. 1e-195;
Matches 1838; Conservative 0; Mismatches 51; Indels 2; Gaps 2;

```

```

QY 3391 aaattaaagataatttaatttagaagaagaanaaaatlcatctactatcaagtgtgata 3450
DB |||||||
DB aaattaaagataatttaatttagaagaagaanaaaatlcatctactatcaagtgtgata 426
QY 3451 catcatttaattgtctgaattaaagaagtaataaaaaataaaattatccagtgtaattc 3510
DB |||||||
DB catcatttaattgtctgaattaaagaagtaataaaaaataaaattatccagtgtaattc 486
QY 3511 ccaagtgaaataatcatcggaattttaaacaatgtcattgaatcttaacaaaaatttcacca 3570
DB |||||||
DB ccaagtgaaataatcatcggaattttaaacaatgtcattgaatcttaacaaaaatttcacca 546
QY 3571 gaagagacagatgttgcacaagttgtgaagtgaagtgatccgcacacattagaacaagt 3630
DB |||||||
DB gaagagacagatgttgcacaagttgtgaagtgaagtgatccgcacacattagaacaagt 606
QY 3631 caaccaagaagaacagcatcactcatgttagagcagagcttaacacaataacaacatca 3690
DB |||||||
DB caaccaagaagaacagcatcactcatgttagagcagagcttaacacaataacaacatca 666
QY 3691 caaaatgtcagatgaattagtagcagtagtaacagtagtaacagtagtaacagtagtaac 3750
DB |||||||
DB caaaatgtcagatgaattagtagcagtagtaacagtagtaacagtagtaacagtagtaac 726
QY 3751 gaagattatgatatttagagcagtagtaacagtagtaacagtagtaacagtagtaac 3810
DB |||||||
DB gaagattatgatatttagagcagtagtaacagtagtaacagtagtaacagtagtaac 786
QY 3811 gatacatccttctaanaattgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 3870
DB |||||||
DB gatacatccttctaanaattgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 846
QY 3871 ggtgtttagaagattttaaacaacattagaagaagaagaagaagaagaagaagaagaaga 3930
DB |||||||
DB ggtgtttagaagattttaaacaacattagaagaagaagaagaagaagaagaagaagaaga 906
QY 3931 gttaaagataattttaaattcacagatttaataaagcgtgaanaatttcaaaatgttttaga 3990
DB |||||||
DB gttaaagataattttaaattcacagatttaataaagcgtgaanaatttcaaaatgttttaga 966
QY 3991 tcagatttaattccatataaagaatttaacatcaagaattatgtgtcgaagatccatatt 4050
DB |||||||
DB tcagatttaattccatataaagaatttaacatcaagaattatgtgtcgaagatccatatt 1026
QY 4051 aaatttctaataaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 4110
DB |||||||
DB aaatttctaataaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1086
QY 4111 tcaatagatcagatataaatttgcgaatgttctgtgatalatatataaattatcc 4170
DB |||||||
DB tcaatagatcagatataaatttgcgaatgttctgtgatalatatataaattatcc 1146
QY 4171 gaaatatataaattcagatttagatccaatttaaaatatatacaacagacaagaagtga 4230
DB |||||||
DB gaaatatataaattcagatttagatccaatttaaaatatatacaacagacaagaagtga 1206
QY 4231 aatgagaataaccccttctttaaacaatgtgaacattatataaacaagcttgaatt 4290
DB |||||||
DB aatgagaataaccccttctttaaacaatgtgaacattatataaacaagcttgaatt 1266
QY 4291 aaatgtattatttgaattcatcttagaagcaaaagctctaattatatacatatgaa 4350
DB |||||||
DB aaatgtattatttgaattcatcttagaagcaaaagctctaattatatacatatgaa 1326
QY 4351 tcaaacgtgaagtttaaaataaagaacttaattacttaaaaaaatccaagaacaattg 4410
DB |||||||
DB tcaaacgtgaagtttaaaataaagaacttaattacttaaaaaaatccaagaacaattg 1386
QY 4411 gcagattttaaaaaataatacaatttcgttgaattgcgatttatacaacaattataac 4470
DB |||||||
DB gcagattttaaaaaataatacaatttcgttgaattgcgatttatacaacaattataac 1446
QY 4471 cataataacttattgacaagttccttagtacagtagtgttttgaanaattgtgtaa 4530

```

```

Db 1447 cataataactatgtgcaagttccttagtgcaggtatgtgttttgaataactctgttaa 1506
    |||
QY 4531 accgttttacttaattacttgatggaacatgcaaggtatgttaaacattcacaac 4590
    |||
Db 1507 accgttttacttaattacttgatggaacatgcaaggtatgttaaacattcacaac 1566
    |||
QY 4591 caatgtgtaaaaaaacaatgctcacaacaaattctgatttttcgacattagtgtaaga 4650
    |||
Db 1567 caatgtgtaaaaaaacaatgctcacaacaaattctgatttttcgacattagtgtaaga 1626
    |||
QY 4651 gaagaatgttaaatgtttatataattacaacaaagaaggtgataaagtgttgtaaaccca 4710
    |||
Db 1627 gaagaatgttaaatgtttatataattacaacaaagaaggtgataaagtgttgtaaaccca 1686
    |||
QY 4711 aatcctacttgtaagaaataatggtgagatgtgacagatgcagaatgtaccgaaga 4770
    |||
Db 1687 aatcctacttgtaagaaataatggtgagatgtgacagatgcagaatgtaccgaaga 1746
    |||
QY 4771 gattcaggtagcaagcaagaaatcacatgtggaatgtacttaacctgattcttcca 4830
    |||
Db 1747 gattcaggtagcaagcaagaaatcacatgtggaatgtacttaacctgattcttcca 1806
    |||
QY 4831 ctttcgagtgtatcttctgcagttcctcctaactcttaggaatcatcttatt-aat 4889
    |||
Db 1807 ctttcgagtgtatcttctgcagttcctcctaactcttaggaatcatcttattaaat 1866
    |||
QY 4890 actcatgttaatatattacaagttcattaa 4920
    |||
Db 1867 actcatgttaatatattacaagttcattaa 1897
    |||

```

RESULT 7

AAT97957
ID AAT97957 standard; DNA; 1896 BP.

AC AAT97957;

XX 03-APR-1998 (first entry)

DE Chimeric MSA-1 antigenic protein 2 used in a malaria vaccine.

KW Signal peptide; malaria; vaccine; merozoite surface antigen-1 peptide;

KW MSA-1; recombinant vaccinia virus; Plasmodium falciparum; anchor peptide;

KW immune response; humoral; cell-mediated; merozoite; ss.

XX Chimeric - Mammalia.

OS Chimeric - Plasmodium falciparum.

XX Key Location/Qualifiers

FT misc_feature 1..165

FT /tag- a

FT /note= "Signal sequence added to improve

FT misc_feature 166..1896

FT /tag- b

FT /note= "Plasmodium falciparum MSA-1 peptide"

XX WO9726911-A1.

XX 31-JUL-1997.

XX 29-JAN-1997; 97WO-US01395.

XX 29-JAN-1996; 96US-0593006.

XX (GEOU) UNIV GEORGETOWN.

XX Davidson EA, Yang S;

XX WPI; 1997-393372/36.

XX Malaria vaccine - comprises expression vector expressing fragment of

PT merozoite surface antigen
XX
PS Claim 33; Fig 3; 75pp; English.
XX
CC The present sequence encodes a chimeric protein that contains a mammalian
CC signal sequence, and a merozoite surface antigen-1 (MSA-1) carboxy
CC terminal peptide. The C-terminal fragment of MSA-1 provides a more
CC specific response than the complete MSA-1, and attachment of the
CC signal sequence improve the immunogenicity of the protein better than
CC the use of an adjuvant. The chimeric protein, and expression vectors
CC (analogues that express MSA-1 without either signal or anchor peptides),
CC particularly in the form of recombinant vaccinia virus, are used in
CC vaccines to prevent or treat malaria caused by Plasmodium falciparum. The
CC vaccinia vector expresses the antigen fragment for many days, or even
CC years, generating a long-lasting immune response (humoral and/or
CC cell-mediated) against the merozoite form of the parasite. In humans or
CC other animals.
XX
SQ Sequence 1896 BP; 813 A; 230 C; 272 G; 581 T; 0 other;

Query Match 35.4%; Score 1741.2; DB 18; Length 1896;
Best Local Similarity 95.3%; Pred. No. 1.1e-190;
Matches 1805; Conservative 0; Mismatches 88; Indels 1; Gaps 1;

```

QY 2972 aagaagatataataataaaaaaaacttacagttacatcattgtattatataataat 3031
    |||
Db 1 atgaagatcatattcttcttattgttcatcttcttcttattataataacaaatgtgta 60
    |||
QY 3032 ataataaataatagaagaattatttgataaaaaagaacagttgtaataataaatgc 3091
    |||
Db 61 aacatgaaagttatatacaagaactgtcacaacaaactgaagctttagaagatgagatgtg 120
    |||
QY 3092 aaattaaaaacttacttattt-aaagaacaaatgaatcaaatgaattcaattaat 3150
    |||
Db 121 acagttatagttatttccaagaagaataatgtataatgaattgaattcaattaat 180
    |||
QY 3151 aaccagaagatattatacaaaacttttcgtttctttaaacaataaaagaagcttga 3210
    |||
Db 181 aaccagaagatattatacaaaacttttcgtttctttaaacaataaaagaagcttga 240
    |||
QY 3211 atagcagaactgaaacacacatagagaacacacaaatattatttgaacattaaagga 3270
    |||
Db 241 atagcagaacactgaaacacacatagagaacacacaaatattatttgaacattaaagga 300
    |||
QY 3271 ctgtttaatatataatggttgatcatcctcattaaaaaacttaagtgaagaatcaat 3330
    |||
Db 301 ctgtttaatatataatggttgatcatcctcattaaaaaacttaagtgaagaatcaat 360
    |||
QY 3331 caacagaagatattatgccaatttagaanaaactttaaagttatgaattagaaga 3390
    |||
Db 361 caacagaagatattatgccaatttagaanaaactttaaagttatgaattagaaga 420
    |||
QY 3391 aaattaaagataatttaatttagaanaagaanaaaattatcatcattatcaagtatt 3450
    |||
Db 421 aaattaaagataattttaaatttagaanaagaanaaaattatcatcattatcaagtatt 480
    |||
QY 3451 catcatttaattgtcgaattaaagaagaatataaaaaataaaattatcacagttact 3510
    |||
Db 481 catcatttaattgtcgaattaaagaagaatataaaaaataaaattatcacagttact 540
    |||
QY 3511 ccaagtgaaataatgagatgttacaatgcatcttagaattcttaacaaaattctccca 3570
    |||
Db 541 ccaagtgaaataatgagatgttacaatgcatcttagaattcttaacaaaattctccca 600
    |||
QY 3571 gaaggaacagatgttgcaacagttgtaagtgaaggtgacccagacattagaacaagt 3630
    |||
Db 601 gaaggaacagatgttgcaacagttgtaagtgaaggtgacccagacattagaacaagt 660
    |||
QY 3631 caaccaaagaacacagcatcactatgttagagcagagttcaacaaataacaacatca 3690
    |||
Db 661 caaccaaagaacacagcatcactatgttagagcagagttcaacaaataacaacatca 720
    |||

```

```

QY 3691 caaatgctgatacgagtagatgacgtaatacatagtaacctatatttggagaatccgaa 3750
    |||
Db 721 caaatgctgatacgagtagatgacgtaatacatagtaacctatatttggagaatccgaa 780
QY 3751 gaagattatgatacttagagcaagtagtaacggggaagcagtaactcttcctcgtaatt 3810
    |||
Db 781 gaagattatgatacttagagcaagtagtaacggggaagcagtaactcttcctcgtaatt 840
QY 3811 gataacatacttctaataatgaaatgtaataatgtaagtttataatttaaacctttagca 3870
    |||
Db 841 gataacatacttctaataatgaaatgtaataatgtaagtttataatttaaacctttagca 900
QY 3871 ggtgtttatagaagtttaaaaaaacaaatgagaataacgtttatgacatttaattgtaatt 3930
    |||
Db 901 ggtgtttatagaagtttaaaaaaacaaatgagaataacgtttatgacatttaattgtaatt 960
QY 3931 gtttaagatattttaaatttcaatgatttaataacgttgaataatttcaaaaatgttttagaa 3990
    |||
Db 961 gtttaagatattttaaatttcaatgatttaataacgttgaataatttcaaaaatgttttagaa 1020
QY 3991 tcagatttaattccataaagaatttaacatacgaatgaattgttgcgaagttccat 4050
    |||
Db 1021 tcagatttaattccataaagaatttaacatacgaatgaattgttgcgaagttccat 1080
QY 4051 aaatttcttaataaagaaaaagagataaattcttaagcagttataattatataagat 4110
    |||
Db 1081 aaatttcttaataaagaaaaagagataaattcttaagcagttataattatataagat 1140
QY 4111 tcaatgatacgataataatttgcgaatgagttcttggatattatataaataattatcc 4170
    |||
Db 1141 tcaatgatacgataataatttgcgaatgagttcttggatattatataaataattatcc 1200
QY 4171 gaaaatataatacagatttagattcaatttaaaaataatacaacgaacgaagtgaa 4230
    |||
Db 1201 gaaaatataatacagatttagattcaatttaaaaataatacaacgaacgaagtgaa 1260
QY 4231 aatgagaataacacctcttctttaaacaalatgagacactataaaacagttatgat 4290
    |||
Db 1261 aatgagaataacacctcttctttaaacaalatgagacactataaaacagttatgat 1320
QY 4291 aaaaattgatttttgaattcattttagaagcaaaagtcttaataatacatgttgaaaa 4350
    |||
Db 1321 aaaaattgatttttgaattcattttagaagcaaaagtcttaataatacatgttgaaaa 1380
QY 4351 tcaaacgtagaagttaaaaataaagaacttaattacttaaaaaacaattcaagaacaaatg 4410
    |||
Db 1381 tcaaacgtagaagttaaaaataaagaacttaattacttaaaaaacaattcaagaacaaatg 1440
QY 4411 gcagattttaaataaatacaaatctcggtggaattgctgattatcaacagattataac 4470
    |||
Db 1441 gcagattttaaataaatacaaatctcggtggaattgctgattatcaacagattataac 1500
QY 4471 cataataactatgtaaaagttccttagtaagagratagtttttgaaaaactgtctaaa 4530
    |||
Db 1501 cataataactatgtaaaagttccttagtaagagratagtttttgaaaaactgtctaaa 1560
QY 4531 accgttttacttaattacttgaatggaacttgcgaagttgtttaaacaatttcaaacac 4590
    |||
Db 1561 accgttttacttaattacttgaatggaacttgcgaagttgtttaaacaatttcaaacac 1620
QY 4591 caatgcgttaaaaaacaatgctccacaataatctcgatglttcagacatttagagaaga 4650
    |||
Db 1621 caatgcgttaaaaaacaatgctccacaataatctcgatglttcagacatttagagaaga 1680
QY 4651 gaagaaatgtaaatggtttatattaataatcaacaagaaggtatataagttgttgaatacca 4710
    |||
Db 1681 gaagaaatgtaaatggtttatattaataatcaacaagaaggtatataagttgttgaatacca 1740
QY 4711 aatcccaactgttaacgaataataatgtgtgagtcgaatcccaatcccaatccgaagaa 4770
    |||
Db 1741 aatcccaactgttaacgaataataatgtgtgagtcgaatcccaatcccaatccgaagaa 1800
QY 4771 gattcagtagaacaaggaaagaataatcaatgtgaaatgtaactaaacctgattctatcca 4830

```

```

Db 1801 gattcagtagcaacggaagaataatcaatgtgaatgtaacttaactgattctatcca 1860
QY 4831 ctcttcagtagtaatttcttcgagttccttaact 4864
Db 1861 ctcttcagtagtaatttcttcgagttccttaact 1894

RESULT 8
AAK56009
ID AAK56009 standard; cDNA; 1088 BP.
XX
AC AAK56009;
XX
DT 31-AUG-1999 (first entry)
XX
DE Merozoite surface protein MSP-1-42 cDNA.
XX
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal; ss.
XX
OS Plasmodium falciparum.
XX
FH Key
FT Location/Qualifiers
FT CDS 1..1086
FT /*tag= a
XX
PN MO920774-A2.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22226.
XX
PR 15-MAY-1998; 98US-0085649.
XX
PR 20-OCT-1997; 97US-0062592.
XX
PA (GEN2 ) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
PI WPI; 1999-288313/24.
XX
DR P-PSDB; AAY09373.
XX
PT Modified malarial protein for use in anti-malarial vaccines
XX
PS Example; Fig 2; 35pp; English.
XX
CC This nucleic acid encodes a 42 kDa C-terminal portion (see AAY09373)
CC of malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The 3' end of the sequence has been
CC modified to include a 6xhis tag. The nucleic acid sequence has
CC been modified (see AAK56008) according to a method of the invention
CC in order to improve expression in mammalian host cells and in
CC transgenic animals. In the modified coding sequence, 306 nucleotide
CC positions have been replaced to lower the AT content (from 76 to
CC 49.7%) and to eliminate 10 mRNA instability motifs. The encoded
CC amino acid sequence is unaltered. In another modified sequence
CC (see AAK56021), a signal peptide sequence has been added and two
CC N-glycosylation sites eliminated. The invention allows expression
CC of MSP-1-42 protein in the milk of transgenic animals, and also
CC provides a DNA vaccine comprising a vector containing the altered
CC MSP-1-42 sequence.
XX
SQ Sequence 1088 BP; 454 A; 139 C; 150 G; 345 T; 0 other:

```

Query Match 21.6%; Score 1063.4; DB 20; Length 1088;
Best local Similarity 99.9%; Pred. No. 1.4e-113;
Matches 1064; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3790 gcagtaactcttcgtaattgataacatacttctaataatgaaatgataatgagtt 3849

```

|||||
Db 1 ggaagaaaccccttcgtaattgataacacattcctaattgaaatgaatagacgggtc 60
QY 3850 ttatatttaaacctttagcaggtgttctataagaatttaaaaaaacattagaataac 3909
Db 61 ttatatlttaaacctttagcaggtgttctataagaatttaaaaaaacattagaataac 120
QY 3910 gttatgacatttaagttaagttaagttaagttaagttaagttaagttaagttaagt 3969
Db 121 gttatgacatttaagttaagttaagttaagttaagttaagttaagttaagttaagt 180
QY 3970 aatttcaaaaatgctttagaaccagatttcaatccatataaagaatttaacatcaat 4029
Db 181 aatttcaaaaatgctttagaaccagatttcaatccatataaagaatttaacatcaat 240
QY 4030 tatgtgttcaagaatccatataaatttctataaagaagaagaagaagaagaagaaga 4089
Db 241 tatgtgttcaagaatccatataaatttctataaagaagaagaagaagaagaagaaga 300
QY 4090 agttataattatataagaatccatataaagaagaagaagaagaagaagaagaagaaga 4149
Db 301 agttataattatataagaatccatataaagaagaagaagaagaagaagaagaagaaga 360
QY 4150 ggaattataaataatataatccgaaataataaatacagatttagattcaatataaata 4209
Db 361 ggaattataaataatataatccgaaataataaatacagatttagattcaatataaata 420
QY 4210 atcaacgaaacaaacgaagtggaataatcagatcccttctttaaacaatattgacc 4269
Db 421 atcaacgaaacaaacgaagtggaataatcagatcccttctttaaacaatattgacc 480
QY 4270 ttatataaacagtgtaagtataaattgatttattgttaattcatttagaagaacaaagt 4329
Db 481 ttatataaacagtgtaagtataaattgatttattgttaattcatttagaagaacaaagt 540
QY 4330 ctaaatatatacatatgagaacatcaaacgtagaagttaaataaagaacattatcatta 4389
Db 541 ctaaatatatacatatgagaacatcaaacgtagaagttaaataaagaacattatcatta 600
QY 4390 aaaaacattcaagacaaatctgtagaatttaaaaaaaataaacaattcgttggaattgct 4449
Db 601 aaaaacattcaagacaaatctgtagaatttaaaaaaaataaacaattcgttggaattgct 660
QY 4450 gattatcaacagattatatacattataactattgacaaagttcctttagacaggtatg 4509
Db 661 gattatcaacagattatatacattataactattgacaaagttcctttagacaggtatg 720
QY 4510 gttttgaaaatctgtctataaacgcttttattctaatctttagatggaacttgcaaggt 4569
Db 721 gttttgaaaatctgtctataaacgcttttattctaatctttagatggaacttgcaaggt 780
QY 4570 atgttaaacattcaacacacacacacacacacacacacacacacacacacacacac 4629
Db 781 atgttaaacattcaacacacacacacacacacacacacacacacacacacacacac 840
QY 4630 ttccagacatttagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 4689
Db 841 ttccagacatttagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 900
QY 4690 gataaatgtttagaagaatccaaacccactttagaagaagaagaagaagaagaagaaga 4749
Db 901 gataaatgtttagaagaatccaaacccactttagaagaagaagaagaagaagaagaaga 960
QY 4750 gatgccaaatgtacccaagaagaatccaggtagcaacgcaagaagaagaagaagaaga 4809
Db 961 gatgccaaatgtacccaagaagaatccaggtagcaacgcaagaagaagaagaagaaga 1020
QY 4810 actaaacccgattctatccactttagcagtgatttttcagat 4854
Db 1021 actaaacccgattctatccactttagcagtgatttttcagat 1065

```

```

AA25587
ID AAX25587 strand: cDNA; 1088 BP.
AC AAX25587;
DT 02-AUG-1999 (first entry)
DE Merozoite surface protein MSP-1-42 cDNA.
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal; ss.
OS Plasmodium falciparum.
FH Key Location/Qualifiers
FT CDS 1..1086
FT tag= a
MO9920766-A2.
PD 29-APR-1999.
PF 20-OCT-1998; 98WO-US22225.
PR 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
PA (GENZ ) GENZYME TRANSGENICS CORP.
PI Chen LH, Meade H;
PI Chen LH, Meade H;
DR WPI: 1999-302742/25.
DR P-PSDB; AAY05833.
XX
XX New modified recombinant nucleic acid sequences useful for producing
XX malarial DNA vaccine
XX
XX Disclosure: Fig 2: 43pp; English.
XX
XX This nucleic acid encodes a 42 kDa C-terminal portion (see AAY05833)
XX of malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
XX important target for the development of a vaccine against
XX Plasmodium falciparum. The 3' end of the sequence has been
XX modified to include a 6xHis tag. The nucleic acid sequence has
XX been modified (see AAX25586) according to a method of the invention
XX in order to improve expression in mammalian host cells and in
XX transgenic animals. In the modified coding sequence, 306 nucleotide
XX positions have been replaced to lower the AT content (from 76 to
XX 49.7%) and to eliminate 10 mRNA instability motifs. The encoded
XX amino acid sequence is unaltered. In another modified sequence
XX (see AAX25593), a signal peptide sequence has been added and two
XX N-glycosylation sites eliminated. The invention provides modified
XX recombinant nucleic acid sequences and methods for increasing the
XX mRNA levels and protein expression of proteins that are difficult
XX to express in cell culture systems, mammalian cell culture systems
XX or in transgenic animals. The preferred difficult protein
XX candidates for expression are those derived from lower organisms
XX such as parasites, bacteria and viruses that have DNA coding
XX sequences of high AT content or which have mRNA instability motifs
XX or rare codons relative to the recombinant expression system to be
XX used. The invention allows expression of MSP-1 protein in the milk
XX of transgenic animals, and also provides a DNA vaccine comprising a
XX vector containing the altered MSP-1-42 sequence.
XX
XX Sequence 1088 BP; 454 A; 139 C; 150 G; 345 T; 0 other;

```

```

Query Match 21.6%; Score 1063.4; DB 20; Length 1088;
Best Local Similarity 99.9%; Pred. No. 1.4e-113;
Matches 1064; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3790 gcaagtaaccccttcgtaattgataacacattcctaattgaaatgaatagacgggtc 3849

```

```

|||||
Db      1 gcagtaactcctcgtaattgataacatacttctaanaattgaaatgatacgaagt 60
OY      3850 ttatatttaaacctttagcaggtgttttagaagttaaaaaaacaaatagaataaac 3909
Db      61 ttatatttaaacctttagcaggtgttttagaagttaaaaaaacaaatagaataaac 120
OY      3910 gttatgacatttaattgtaattgtaaggaattttaaatcacgatttaataacgtgaa 3969
Db      121 gttatgacatttaattgtaattgtaaggaattttaaatcacgatttaataacgtgaa 180
OY      3970 aatttcaaaaatgttttagaatacgaatttaattccatataagaatttaacatgaat 4029
Db      181 aatttcaaaaatgttttagaatacgaatttaattccatataagaatttaacatgaat 240
OY      4030 tatgtgtcaaaagatccataataattcctaataagaagaagaagaataattcctaagc 4089
Db      241 tatgtgtcaaaagatccataataattcctaataagaagaagaagaataattcctaagc 300
OY      4090 agttataattatataaggaattcaatagatacgaataaatttcgaatgtgtctt 4149
Db      301 agttataattatataaggaattcaatagatacgaataaatttcgaatgtgtctt 360
OY      4150 ggaatttataaaaatattatccgaaaaatataaatcagatttagatccaatttaaaaat 4209
Db      361 ggaatttataaaaatattatccgaaaaatataaatcagatttagatccaatttaaaaat 420
OY      4210 atcaaacgacaaaaggtgtaaaatgagaatacctccctttttaacaaatattgagac 4269
Db      421 atcaaacgacaaaaggtgtaaaatgagaatacctccctttttaacaaatattgagac 480
OY      4270 ttatataaaacagtttaattgataaaattgattttgtaattttagaagaagaagt 4329
Db      481 ttatataaaacagtttaattgataaaattgattttgtaattttagaagaagaagt 540
OY      4330 ctaaatatatacattgagaataatcaaacgttagaagtttaaaataagaacttaattctta 4389
Db      541 ctaaatatatacattgagaataatcaaacgttagaagtttaaaataagaacttaattctta 600
OY      4390 aaaaacattcaagacaattgtgcagattttaaaaaaatatacaatttcgttgaattgct 4449
Db      601 aaaaacattcaagacaattgtgcagattttaaaaaaatatacaatttcgttgaattgct 660
OY      4450 gatttaacaagattataacccaataaacttaattgacaagaagttcccttagaaggtatg 4509
Db      661 gatttaacaagattataacccaataaacttaattgacaagaagttcccttagaaggtatg 720
OY      4510 gtttttgaataattctgtctaaacggttttaacttaatttactgtgaaactgtcaaggt 4569
Db      721 gtttttgaataattctgtctaaacggttttaacttaatttactgtgaaactgtcaaggt 780
OY      4570 atgttaacaatttcaacaacccaatcgttaaaaaaacaatgtccacaataatttgatgt 4629
Db      781 atgttaacaatttcaacaacccaatcgttaaaaaaacaatgtccacaataatttgatgt 840
OY      4630 ttcaagacatttagatgaagaagaagaatgttaaatgttttaataattacaacaagaaggt 4689
Db      841 ttcaagacatttagatgaagaagaagaatgttaaatgttttaataattacaacaagaaggt 900
OY      4690 gataaatgtgtgtaaaatcccaaatcctactgttaacgaaataatgtgtgagtgtgca 4749
Db      901 gataaatgtgtgtaaaatcccaaatcctactgttaacgaaataatgtgtgagtgtgca 960
OY      4750 gattgtcaaatgtgacgaagaagaatgttagtagaaggaagaagaataatcagtgtatgt 4809
Db      961 gattgtcaaatgtgacgaagaagaatgttagtagaaggaagaagaataatcagtgtatgt 1020
OY      4810 actaaacccgattcttatccactttcgaatgtgattttcgcaggt 4854
Db      1021 actaaacccgattcttatccactttcgaatgtgattttcgcaggt 1065

```

RESULT 10

```

AAC68976
ID      AAC68976 standard; DNA; 1077 BP.
XX
AC      AAC68976.
XX
DT      27-FEB-2001 (first entry)
XX
DE      Merozoite surface protein-142 coding sequence.
XX
KW      Merozoite surface protein; protozoacide; vaccine; malaria; ss.
XX
OS      Plasmodium falciparum.
XX
PN      WO200063245-A2.
PD      26-OCT-2000.
PE      20-APR-2000; 2000MO-GB01558.
PF      XX
PR      20-APR-1999; 99GB-0009072.
PR      13-MAY-1999; 99US-0311817.
PR      25-MAY-1999; 99CA-2271451.
XX
PA      (MED1-) MEDICAL RES COUNCIL.
PI      Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthairibull C;
DR      WPI; 2001-015762/02.
XX
PT      Novel variants of the C-terminal fragment of Plasmodium merozoite
PT      surface protein-1, useful as vaccines for treating or preventing
PT      malaria -
XX
PS      Example 5; Fig 15; 126pp; English.
XX
CC      The present invention relates to non-natural variants of a C-terminal
CC      fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC      non-natural variants have reduced affinity for at least 1 antibody
CC      capable of blocking a second antibody that inhibits the proteolytic
CC      cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
CC      one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC      MSP-1.4.2, compared to natural MSP-1.1-9. The non-natural variants of the
CC      present invention are useful for immunising a mammal against malaria, and
CC      can be used to treat malaria. The present sequence is the MSP-142
CC      coding sequence.
XX
SQ      Sequence 1077 BP; 338 A; 228 C; 225 G; 286 T; 0 other;

Query Match      13.0%; Score 641.4; DB 22; Length 1077;
Best Local Similarity 74.8%; Pred. No. 1.4e-65;
Matches 804; Conservative 0; Mismatches 271; Indels 0; Gaps 0;

OY      3790 gcagtaactcctcgtaattgataacatacttctaanaattgaaatgaataatgaggt 3849
Db      1 gctgttaccatccatctgtatcgataacattcgtctcaagattgagaagaacgaggtc 60
OY      3850 ttatatttaaacctttagcaggtgttttagaagttaaaaaaacaaatagaataaac 3909
Db      61 ttatatttaaacctttagcaggtgttttagaagttaaaaaaacaaatagaataaac 120
OY      3910 gttatgacatttaattgtaattgtaaggaattttaaatcacgatttaataacgtgaa 3969
Db      121 gttatgacatttaattgtaattgtaaggaattttaaatcacgatttaataacgtgaa 180
OY      3970 aatttcaaaaatgttttagaatacgaatttaattccatataagaatttaacatgaat 4029
Db      181 aatttcaaaaatgttttagaatacgaatttaattccatataagaatttaacatgaat 240
OY      4030 tatgtgtcaaaagatccataataattcctaataagaagaagaagaataattcctaagc 4089
Db      241 tatgtgtcaaaagatccataataattcctaataagaagaagaagaataattcctaagc 300

```


PF	20-OCT-1998;	98WO-US22226.
PR	15-MAY-1998;	98US-0085649.
PR	20-OCT-1997;	97US-0062592.
PA	(GENZ)	GENZYME TRANSGENICS CORP.
P1	Chen LH, Meade H;	
DR	WPI: 1999-288313/24.	
XX	P-PSDB: AAY09372.	
PT	Modified malarial protein for use in anti-malarial vaccines	
PS	Claim 1; Fig 1; 35pp; English.	
XX		
CC	This novel, modified nucleic acid encodes the 42 kDa C-terminal	
CC	portion (see AAY09372) of malaria merozoite surface protein MSP-1	
CC	(MSP-1-42), an important target for the development of a vaccine	
CC	against Plasmodium falciparum. The nucleic acid sequence has been	
CC	modified compared to the native sequence of MSP-1-42 (see AAY56009)	
CC	such that 306 nucleotide positions have been replaced to lower the	
CC	AT content (from 76 to 49.7%) and to eliminate 10 mRNA instability	
CC	motifs while maintaining the same protein amino acid sequence.	
CC	These alterations allow MSP-1-42 to be expressed in mammalian cell	
CC	culture and in transgenic mice. Native MSP-1-12 is known to be	
CC	difficult to express in cell culture systems, mammalian cell	
CC	culture systems or in transgenic animals. The invention allows	
CC	expression of MSP-1-42 protein in the milk of transgenic animals,	
CC	and also provides a DNA vaccine comprising a vector containing the	
CC	altered MSP-1-42 sequence.	
XX		
SQ	Sequence 1065 BP: 324 A; 254 C; 280 G; 207 T; 0 other;	
	Query Match	12.2%; Score 601; DB 20; Length 1065;
	Best Local Similarity	72.8%; Pred. No. 5,6e-61;
	Matches 775; Conservative 0; Mismatches 290; Indels 0; Gaps	0;
OY	3730	gcagtaactctctccgtaattgataacatacttctaataatggaatgatgagtt 3849
DB	1	gccgcatactccctccgatacatcctctccaagatcgagaacgagtcgagtg 60
OY	3850	ttaatttaaacctttagcagtgctttagaagttttaaanaaacatgaanaataac 3909
DB	61	ctgacccgaaagcgcctggcaggggtctacgcggagcctgaaagaagctcgagaacac 120
OY	3910	gttaacatttaagtttaattgtaagatatttaattcaattcaactltaataacgtga 3969
DB	121	gtgagtaaccttaacagtgaaagtgaaatcccgaaagcgcgttcaacaagcgagg 180
OY	3970	aattcaaaaatgctttagaatacagatttaattcatalaagaatttaacatcaat 4029
DB	181	aacttcaagaagtgctgagagagcattcgatccctacaagaattctgcagcagcaac 240
OY	4030	tatgtgtcaagaagatcccatataatttcttaataagaanaaaggataattcttagc 4089
DB	241	tacggtgacaaaggatccctacaagtlctcgaaacaaggagagataagtlctcgagc 300
OY	4090	agtttaattatataagatcatcatagatacggatataaatttggcaatggtctt 4149
DB	301	agttacaactacatacaagaatgattgttagatccgatataacttgcgaacgattctcg 360
OY	4150	ggatattataaataatttcgcgaanaalataaalcagatttgatitcaattaaaaat 4209
DB	361	ggatactcaaatccctgcgagaagatacagagcattcgattccaataagaagtac 420
OY	4210	atcaacggcaaaaacagggtgaataatgagaatcccttcccttttaacacatttgagac 4269
DB	421	atcaacgataagcaggagagaaagagaatgactccgcttccgaaacaacatcgagacc 480
OY	4270	ttataataaacggttaatgataaaattgattattgttaattcatttgaagaacaaagt 4329

[illegible]

Dk

07

Dik

Oy

Dk

Ox

Dh

Oy

Dk

Q5

Dh

QY

Dh

Q4

Dh

Q5

Dh

QY

Dh

Q5

Db 703 gccgacatgcagccagattacaaccacaacacctgtcgcacaaagtctctgagcccgaa 762
QY 4507 atggttttaaaatctgtctaaacccgtttatctatctacttgaatggaacttgcaa 4566
Db 763 atggtcttcgaaaacccgtgcgaagaccgctctctgagcaacctgtctgttgaaacctgcag 822
QY 4567 ggtatgttaaacattacaaccaccaatgcgtlaaaaaaacaatgacacaaattctgga 4626
Db 823 ggaatgctgagatcagccagccagcagctgtctggaagaagcagtgctcccgagaacgcgga 882
QY 4627 tgttcgacatttagtctgaagaagaatgtcaaatgtttatataattcaacaaga 4686
Db 883 tgccttcgaacaccgtgagtgaggagtgcaagtgcctgtcgtcaactacaagcaggaa 942
QY 4687 ggtgataaagctgttgaacatcccaatccctactgttaacgaaaataatgtgtgag 4746
Db 943 gggagatagtggttggaataaaccccaatccctactgttaacgaaaataatgtgtgag 1002
QY 4747 gccagatgcacaatgttacggaagaagatcagtagcaacggaagaagaatacatgtgaa 4806
Db 1003 gccagatgcacaatgttacggaagaagatcagtagcaacggaagaagaatacatgtgaa 1062
QY 4807 tgcataaactgtattctatccacttttcgattgtgtatttctgcact 4854
Db 1063 tgcataaactgtattctatccacttttcgattgtgtatttctgcact 1110

RESULT 14

AAK25593
ID AAK25593 standard; cDNA; 1140 BP.

AC AAK25593;

XX 02-AUG-1999 (first entry)

XX Merozoite surface protein MSP-1-42 modified cDNA.

XX MSP-1; merozoite surface protein; malaria; vaccine;

KM protein engineering; protein expression; codon usage;

KM transgenic animal; mutant; ss.

XX Plasmodium falciparum.

OS Synthetic.

XX Key Location/Qualifiers

FT CDS 1..1131

FT sig_peptide 1..45

FT mat_peptide 46..1128

FT /tag- c

PN WO920766-A2.

XX 29-APR-1999.

XX 20-OCT-1998; 98WO-US22225.

XX 15-MAY-1998; 98US-0085649.

XX 20-OCT-1997; 97US-0062592.

XX (GEN2) GENZYME TRANSGENICS CORP.

XX Chen LH, Meade H;

XX WPI, 1999-302742/25.

XX P-PSDB; AAY05834.

XX New modified recombinant nucleic acid sequences useful for producing

XX malarial DNA vaccine

CC This novel, modified nucleic acid encodes a 42 kDa C-terminal
CC part (see AAY05834) of malaria merozoite surface protein MSP-1
CC (MSP-1-42), an important target for the development of a vaccine
CC against Plasmodium falciparum. The nucleic acid sequence has been
CC modified compared to the native sequence of MSP-1-42 (see AAK25587)
CC such that the AT content has been reduced and 10 mRNA instability
CC motifs eliminated while maintaining the protein amino acid sequence.
CC In addition, a sequence encoding a 15-amino acid beta-casein signal
CC peptide has been added to the 5' end of the sequence, and N262Q
CC and N181Q mutations have been introduced to eliminate
CC N-glycosylation sites. These alterations allow MSP-1-42 to be
CC expressed in the mammary gland (i.e. milk) of transgenic mice. The
CC invention provides modified recombinant nucleic acid sequences and
CC methods for increasing the mRNA levels and protein expression of
CC proteins that are difficult to express in cell culture systems,
CC mammalian cell culture systems or in transgenic animals. The
CC preferred difficult protein candidates for expression are those
CC derived from lower organisms such as parasites, bacteria and
CC viruses that have DNA coding sequences of high AT content or which
CC have mRNA instability motifs or rare codons relative to the
CC recombinant expression system to be used. The invention allows
CC expression of MSP-1 in the milk of transgenic animals, and also
CC provides a DNA vaccine comprising a vector containing the altered
CC MSP-1-42 sequence.

XX Sequence 1140 BP; 353 A; 282 C; 290 G; 215 T; 0 other;

XX Query Match 12.0%; Score 588; DB 20; Length 1140;

XX Best Local Similarity 71.9%; Pred. No. 1.7e-59;

XX Matches 768; Conservative 0; Mismatches 300; Indels 0; Gaps 0;

QY 3787 gaagcagtaactctccctccgaattgataacatacttctcaaaattgaatgaatgag 3846
Db 43 gccgcgcacactccctccgcgcacgcataacacatccctgcgaagatgcagagtcag 102
QY 3847 gtttatattaaacactttagcaggtgtttatagaagtttaaaaaacacattagaat 3906
Db 103 gtgtgtacttgagcccccgcagagtgctacagaagccttgagaagagcgcgggaac 162
QY 3907 aacgttagacatttaattgttaattgaagataatttaattcagcagatttaaacgt 3966
Db 163 aacgttagacatttaattgttaattgaagataattcagcagcaggttcaacaagag 222
QY 3967 gaaattcaaaaagtgtttagatcagatcagattatccatataagaattacaatcagt 4026
Db 223 gagaactccaagaacgtgtgcggagcgatctgattccccaagaagatcgcagcagc 282
QY 4027 aattatgttgcagaagatccataaaattcttcaataaagaaaagagataaattctta 4086
Db 283 aactagtggtcaagaatccctcaaaagtcttcaacaagaagagagataagttctctg 342
QY 4087 agcagttataattatataagaattcaatagatcagagataaaatttgcaatgatt 4146
Db 343 agcagttataattatataagaattcaatagatcagagataaaatttgcaatgatt 4202
QY 4147 ctggatattataaattatccgaaaataataaaatcagatgaattcaatttaaaaaa 4206
Db 403 ctggatattataaattatccgaaaataataaaatcagatgaattcaatttaaaaaa 4262
QY 4207 tatacgaacacaagaagtgaaaatgaagaatccctctttaaacaatatgag 4266
Db 463 tatatcaagataagcagaggaagaaagtacacctctctctgaaacaatcagag 522
QY 4267 acctatataaacaagttaaatgaattgatttctgttaattcaatttagaagcaaa 4326
Db 523 acctgtacaagcgcgtacaagataagattgtctgtgtccaccgtggagcgaag 582
QY 4327 gtctaatatatacatatgagaatacaacgtagaagtttaaaataaagaactaatttc 4386
Db 583 gtctgcagtaacacataltgagaagcaacgtgaggtlcaagaatcagaagcgtgaattac 642
QY 4387 ttaaaaacaaltcaagaacaattgcagattttaaaaaataacaattcgttggaatt 4446

```

Db      643  ctgaagaccatccaggatgaagctgcccatttcaagaagaacaaacttcgcggaac 702
Qy      4447  gctgatttatacaacataataacataaacttattgaagaagtctcttaagcaagt 4506
Db      703  gccgactctgagcaccattacaacacaaacttcgtcccaagttcctgaagcaccgga 762
Qy      4507  atggttttgaagaactcttcaaaacggttttacttaattacttgaatgaacttgaac 4566
Db      763  atggttcttgaagaacttcgcaagaacgcttcctctgaagaacttcgtatgtgaacccgca 822
Qy      4567  ggtatgttaaacatttcaacaacccaatgcgttaaaaaaacaatgctcacaaatcttga 4626
Db      823  ggaatctcgaagtcagcagcagcagctgtgtgaagaagaagctcccccagaacagcgga 882
Qy      4627  tgttttaagcatttaagaagaagaagaatgttaaatgttttataattacaacaaga 4686
Db      883  tgcctcagaacacttgatgagagaggtgcaagtgccctgctgaactacaagcagga 942
Qy      4687  ggtgaacaatgtgttaaaatcccaatccctactgtgaacgaataatgtgtgagtgtat 4746
Db      943  ggaagataagtgtgtgaaaaaaccacactctactgttaacgagaacatgagagtgcat 1002
Qy      4747  gcaagatgcgaatgtacccgaagaagaatcaggtagcaacggaagaagaacatcatgtga 4806
Db      1003  gccgactgcgaagtgtacccgagagagatcaggaagcaacggaagaagaatcacctgcga 1062
Qy      4807  tgtactaaacccgtatcttaccacttccgatttcgatttcgcagt 4854
Db      1063  tgtaccaagcctgattctatccacttcgatttcgatttcgcagt 1110

```

RESULT 15

```

AA81151
ID      AA81151 standard; DNA; 1068 BP.
XX
AC      AA81151;
XX
DT      11-NOV-1990 (first entry)
XX
DE      DNA sequence encoding polypeptide p190-3.
XX
KW      Polypeptide p190-3; P.falciparum; merozoite; vaccine.
XX
OS      Plasmodium falciparum.
XX
FH      Key      Location/Qualifiers
FT      CDS      1..1068.
FT      CDS      /*tag= a
XX
PN      EP283829-A.
XX
PD      28-SEP-1988.
XX
PF      08-MAR-1988; 88EP-0103564.
XX
PR      19-MAR-1987; 87GB-0006599.
XX
PA      (HOFF) F. HOFFMANN-LA ROCHE & CO.
XX
PI      Ulrich C, Gentz H, Takacs B;
XX
DR      MPI: 1988-272339/39.
DR      P-PSDB; AAP80547.
XX
PT      New immunogenic polypeptides - derived from Plasmodium falciparum 190 kD
XX      surface antigen precursor, useful in vaccines, and encoding DNA sequences
XX
PS      Disclosure; 5 spp. German.
XX
CC      The DNA encodes a region of a polypeptide. The region contains at least
CC      one epitope of the 190kD precursor of the major merozoite surface antigen
CC      of P.falciparum.

```

```

XX
SQ      Sequence 1068 BP; 471 A; 129 C; 161 G; 307 T; 0 other;
Query Match      10.9%; Score 535; DB 9; Length 1068;
Best Local Similarity 87.9%; Pred. No. 1.8e-53;
Matches 583; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
Qy      331  acaatccctcaagatattcaagtgatcagatcgaactcaacttcaagctgatttaaacac 390
Db      298  ataaaaataaaattatataaggtlaattcccaagcgttaataatatacggattgaacaal 357
Qy      391  agagtcgaacttactgttcaactcaagaagaactcaaatatcccaactcttgattta 450
Db      358  gcatatgaatcttcaaaaaaattcccccagaagagacagatgttgcacagttgtaagt 417.
Qy      451  actaatcatgtttaaacttctgtgtgaabaatcatcagtttcaaatatttaattgata 510
Db      418  gaaagtgcacccgaacttctgtgtgaataattcatcagtttcaaatatttaattgata 477
Qy      511  tatgaagaattatgaattatataataataaactttattttattataagaaga 570
Db      478  tatgaagaattatgaattatataataataaactttattttattataagaaga 537
Qy      571  aaataatgatgtatgtgttaattgattatgttcaaatcccttcaactttaaattcgt 630
Db      538  aaatataatgatgtatgtgttaattgattatgttcaaatcccttcaactttaaattcgt 597
Qy      631  gcaatgaattagacgtacttaaaaaactgtgttcggatatagaagaacattagaacaal 690
Db      598  gcaatgaattagacgtacttaaaaaactgtgttcggatatagaagaacattagaacaal 657
Qy      691  attaagaatattagaaaaaagtggaagattacatttaaaaaaataaaaaaccataga 750
Db      658  attaagaatattagaaaaaagtggaagattacatttaaaaaaataaaaaaccataga 717
Qy      751  aatataatgaatttaattgagaagaatgaagaacaattgataaaaaataagaatgcaact 810
Db      718  aatataatgaatttaattgagaagaatgaagaacaattgataaaaaataagaatgcaact 777
Qy      811  aaagaagaagaaaaaaaataatataccaagctcaatatatgacttcttataataaa 870
Db      778  aaagaagaagaaaaaaaataatataccaagctcaatatatgacttcttataataaa 837
Qy      871  caattgaagaagacacataatttaataagcgttttgaaaaaagcttgaacactttaa 930
Db      838  caattgaagaagacacataatttaataagcgttttgaaaaaagcttgaacactttaa 897
Qy      931  aaaaatgaacaacttaaggaattacttgaataaataatgaataaaatcccccacg 990
Db      898  aaaaatgaacaacttaaggaattacttgaataaataatgaataaaatcccccacg 957
Qy      991  gcc 993
Db      958  gcc 960

```

Search completed: October 27, 2001, 15:51:48
 Job time: 18444 sec

Run on: October 27, 2001, 09:19:24 ; Search time 8587.58 Seconds
(without alignments)

8861.789 million cell updates/sec

Title: US-09-269-874-1
Perfect score: 4920
Sequence: 1 atgaagatcatactctttt.....tattatacagtttcatttaa 4920

Scoring table: IDENTITY_NUC
Gapop. 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
```

Database :

1:	gb_ba1.*
2:	gb_ba2.*
3:	gb_ba3.*
4:	gb_in1.*
5:	gb_in2.*
6:	gb_in3.*
7:	gb_ov.*
8:	gb_ov.*
9:	gb_pat1.*
10:	gb_pat2.*
11:	gb_ph.*
12:	gb_p1.*
13:	gb_p12.*
14:	gb_p13.*
15:	gb_p14.*
16:	em_ba1.*
17:	em_ba2.*
18:	em_fun.*
19:	em_hqo_hum.*
20:	em_hqo_inv.*
21:	em_hqo_rtd.*
22:	em_hqo_hun1.*
23:	em_hqo_hun2.*
24:	em_hqo_hun3.*
25:	em_hqo_hun4.*
26:	em_hqo_hun5.*
27:	em_hqo_hun6.*
28:	em_hqo_hun7.*
29:	em_hqo_hun8.*
30:	em_hqo_inv1.*
31:	em_hqo_inv2.*
32:	em_hqo_rtd.*
33:	em_hun1.*
34:	em_hun2.*
35:	em_hun3.*
36:	em_hun4.*
37:	em_hun5.*
38:	em_hun6.*
39:	em_hun7.*
40:	em_hun7.*
41:	em_in.*
42:	em_or.*
43:	em_or.*

44:	em.ov:
45:	em.ph:
46:	em.ph:
47:	em.ro:
48:	em.ro:
49:	em.sts:
50:	em.sts:
51:	em.sts:
52:	em.vi:
53:	gd.sts:
54:	gd.sts:
55:	gd.sts:
56:	gd.sts:
57:	gd.sts:
58:	gd.vi:
59:	gd.vi:
60:	gd.vi:
61:	gd.vi:
62:	gd.vi:
63:	gd.vi:
64:	gd.vi:
65:	gd.vi:
66:	gd.vi:
67:	gd.vi:
68:	gd.vi:
69:	gd.vi:
70:	gd.vi:
71:	gd.vi:
72:	gd.vi:
73:	gd.vi:
74:	gd.vi:
75:	gd.vi:
76:	gd.vi:
77:	gd.vi:
78:	gd.vi:
79:	gd.vi:
80:	gd.vi:
81:	gd.vi:
82:	gd.vi:
83:	gd.vi:
84:	gd.vi:
85:	gd.vi:
86:	gd.vi:
87:	gd.vi:
88:	gd.vi:
89:	gd.vi:
90:	gd.vi:
91:	gd.vi:
92:	gd.vi:
93:	gd.vi:
94:	gd.vi:
95:	gd.vi:
96:	gd.vi:
97:	gd.vi:
98:	gd.vi:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4920	100.0	4920	9	A92450	A92450 Sequence 1
2	4920	100.0	5917	96	PF09119	X02919 Plasmodium
3	4920	100.0	7038	96	PF09195A	X15063 Plasmodium
4	4779.2	97.1	5760	9	A04562	P. falciparu
5	4779.2	97.1	5760	10	E00656	E00656 CDNA encodi
6	4750.4	96.6	5282	96	PF0190	X03371 P. falciparu
7	2853.4	58.0	5271	5	AF062349	AF062349 Plasmodiu
8	2814.4	57.2	4940	9	A92451	A92451 Sequence 2

[illegible]

ALIGNMENTS

RESULT	1
A92450	
LOCUS	A92450
DEFINITION	Sequence 1 from Patent WO9814583.
ACCESSION	A92450
VERSION	A92450.1
KEYWORDS	GI:6741180
SOURCE	unidentified. unclassified.
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 4920)
AUTHORS	Pan,W. and Bujard,H
TITLE	METHOD FOR PRODUCING RECOMBINANTS INTENDED FOR USE IN A COMPL
JOURNAL	MALARIA ANTIGENE GP190/MSP1 Patent: WO 9814583-A 1 09-APR-1998; PAN WEIDONG (DE); BUJARD HERMANN (DE) Location/Qualifiers 1..4920
FEATURES	/organism="unidentified" /db_xref="taxon:32644"
source	
BASE COUNT	2194 a 589 c 679 g 1458 t
ORIGIN	

Query Match	100.0%	Score 4920;	DB 9;	Length 4920;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 4920;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY 1 atgaagatcatattcttttaagtcaattctttttattataaatacacaaatgta 60
|||||
Db 1 ATGAGATCATATTCTTTTANGTCAATTCCTTTTATATATAAATACCAACATGTGA 60

Qy	61	accatgaagaattacaagaactctgcaaaaaacacgaagctttagaagaatgcagattg	120
Db	61	ACACATGAAAGATTATCAAGACTGTCAAAAAACAATGAAAGCTTTAAGAAAGCAGTATTG	120
Qy	121	acaggtatagattatctcaaaagaaaaatggtattaaatgaaggaacgaaygaca	180
Db	121	ACAGGTTAATAGTTTATTTCAAAAGAAAAAATGTAATTAATGAAGAACAAAGTGGACA	180
Qy	181	gcgtttacaactagtaacccttggttcaaaaggtttcagttcgaatggttcaggtgagc	240
Db	181	GCGTTTACAACTATGTAACCTTGTTCAAAAGGTTACAGTTCAGTTCAGGTGCGC	240
Qy	241	tcaagttgcttcaggtgagtcagttgcttcaggtgagtcagttgcttcaggtgagtc	300
Db	241	TCAGTTGCTTCAGGTGCGTCAGTTGCTTCAGGTGCGTCAGTTGCTTCAGGTGCGTCAGTT	300
Qy	301	gcttcaggtgaggttcaggttaattccaagaagtcacaatccctcagaataatccaagtgatca	360
Db	301	GCTTCAGGTGCTTCAGGTGAATTCAGAGCTCAAAATCCCTCAGTAATTCACAGTGAATCA	360
Qy	361	gagtcataaattccaaagtcgatttaaaacacagagtcagaatlaactggttaactatcaaa	420
Db	361	GATGCTAATATCTTACGCTGATTTAAACACAGAGTACGAAATTACTTGACTATACAAA	420
Qy	421	gaacctcaaatatccccaactcttgatttaactaatcatalgthaacttgtygataat	480
Db	421	GAACCTCAAAATATCCCAACTCTTGATTTAACTAATCATATGTAATGTTTGTGTGAAT	480
Qy	481	attcagtggttcaaatatttaattgagatagaagaatlaatgaattatataataa	540
Db	481	ATTCAATGCTTCAAAATATTTAATGATGAGATATCAAGAAATTAATGCAATTAATTAATAA	540
Qy	541	ttaaaccttatttgatttaataagaagcaaaatlaaigtatglatgltcaatgatac	600
Db	541	TTAAACTTTTATTTTGATTTATTAAGAGCAAAATTAATATGATGATGCTCATATGATAT	600
Qy	601	tgtcaaatcacttccaatctctaaaatctgycgaatgaatgaacttaacttaaaaaact	660
Db	601	TGTCAAAATACCTTTCAATCTTAAATTTGTCGAATGAATTAACGCTACTTAAAAACTT	660
Qy	661	gtgttcggaatagaaaaaatagacacatatlaaagaatlaagyaanaatggaagat	720
Db	661	GTTTCGGATATTAACAAAACCATTTAGCAATTAATTAAGATTAATGTACGAAAAATGAGAT	720
Qy	721	tacattaaaaaaataaaaaaacatagaagaatataatgaatttaattgaagaagtaag	780
Db	721	TACATTTAAAAAATTAAAAAACCATTAAGAAATTAATTAATTAATTAATTAATTAAG	780
Qy	781	aaaaaatctgaataaataaagaatgcaactcaagaagaagaaaaaaatataaccaa	840
Db	781	AAAACAATTTGATTAATAATTAAGATGCACTTAAGAGAAAGAAAAAATAATATACAA	840
Qy	841	gtcaatatgatctcttctatttacaataaacaattagaagaagcacaatttaataagc	900
Db	841	GCTAATATGATCTCTTCTATTATTACATTAACAATTAAGAAAGCACTAATTTAATTAAGC	900
Qy	901	gttttagaaaaacgtlatgcaacttcaaaaaaaatggaacataaagaatlaactgcat	960
Db	901	GTTTTCAGAAAACGTAATGACACTTTAAAAAATAATGAACATTAATGAAGATTAATCTTGAT	960
Qy	961	aaggtataatgaataataaaatccccaccggccgaactctggaataatacccaatlaacttc	1020
Db	961	AAGGTAAATTAATTAATAATCCCCACCGGCCCAATTTGCAAAATATACCAAAATATCTTC	1020
Qy	1021	cttgataagaacaaaaaaatcgaagaacgcggaaaaaaatataaagaattgccaact	1080
Db	1021	CTTGATTAAGACAAAAATTCGAGGACACGAAAAAGAAATTAAGAAATTTGCCAAAAC	1080
Qy	1081	attaaatlaataatlgatagttatlaactgtaaccaactggaatagaaataactatlaaga	1140
Db	1081	ATTAATTTAATTAATGATAGTTTATTTACTGATCCACTTGAATTAATGAATTAATTAAGA	1140


```
OY 1141 gaaaaataaaaaatactgatatagtgcaaggttgaaacaaaggaatcaatgaacc 1200
|||||
DB 1141 GAAAAAATATTAATTTGATATTAAGTGCAGAGTTGMAAGCAAGATCAATGAAACC 1200
OY 1201 aatgaatcccaatgagatctacttaaccttgctacataagatactaaatgcttca 1260
|||||
DB 1201 AATGAATATCAAAATGAGACTTACTTATCCCTTGTATATTAACGATATTAACATGCTTTA 1260
OY 1261 aatgaatcaatctcttggttgatcttaataccaatcttgatatacaaaagcaagc 1320
|||||
DB 1261 AATGAATCAATTTCTTGTGATTAATTAATCAATTTGATTAATCAAAAGAACCAAGT 1320
OY 1321 aaaaaataatactgatatgaaagaaaaaaattcataaatagaattgaagaaaaatt 1380
|||||
DB 1321 AAAAAATATATCTCATATTAAGCAAAAAATTTCTATTAATGAATTAAGCAAAAAATTT 1380
OY 1381 aaaaatagaaaaaaaataatctgaatctgataaaaaaactctacgaagacagatcaagct 1440
|||||
DB 1381 AAAATAGAAAAAATAATTTGAATCTGATTAATAAATCTTACGAGACAGATCTAAGTCT 1440
OY 1441 ttaaatgataatacaaaagaaatagaaataatcttaatagaattatagatacaatct 1500
|||||
DB 1441 TTAATAGATATTAACAAAGATTAATGAATAATTTCTAATGAATTAATGAATCAAAATTC 1500
OY 1501 aataataataagatttaactaattcgaaaaaatgagtggaagatactcataaa 1560
|||||
DB 1501 AATTAATATATGATTTAACTTATTCGAAAAATGATGGGTAAAGATTTCTATATTA 1560
OY 1561 gttgaagaactacacacataaactcttgacacacacacacacacacacacacacacac 1620
|||||
DB 1561 GTGAGAACTTACACACCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1620
OY 1621 gaaaaatcaaaaaactcttaataataatagaaagatatactcttaaggaatagtgat 1680
|||||
DB 1621 GAAAAATTAACAAAGCTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1680
OY 1681 gaaaaaatcaaaatcataataaaatttaataagcaaaatagaaatgagatgaaaca 1740
|||||
DB 1681 GAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1740
OY 1741 ttagttgaaaaatactaaagaatgaaagacgctcttgaaaaaaaacttaactaaagac 1800
|||||
DB 1741 TTAGTTGAAAAATTTTAAAAAGATGAAGAACACCTTTTAAAAAATAATTAATTAATTA 1800
OY 1801 gaaaaataaaccaatgaaaaaatttgaagaatctgacacacacacacacacacacacac 1860
|||||
DB 1801 GAAAAATTAACCAAGATGAAGAAATTTTGAAGATCTGACATTTGTAAGATTAAGATTTCA 1860
OY 1861 aaagtcttatctgaacaaacttgacgaatlaaaaaagacacacacacacacacacacac 1920
|||||
DB 1861 AAAGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1920
OY 1921 gtagaatlaaacaataatcatatgctcccaatctcttaacaacaagaataaagcaaga 1980
|||||
DB 1921 GTAGAAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1980
OY 1981 ccttatattatcttggttgaaaaaagaaatgataaatttaaaagtggtcattgacaaag 2040
|||||
DB 1981 CCTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2040
OY 2041 gtagaatcatctgaataatgagaagaaaaaaacataaaacagaaggtcaatcgataat 2100
|||||
DB 2041 GTAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2100
OY 2101 tcggaacacataacacgaaggaataaacaagacacacacacacacacacacacacacac 2160
|||||
DB 2101 TCGGAACATTAACCAAGCAAGAAATTAACAGCAACCAATTAACCAACCTCGGACCAACAA 2160
OY 2161 gacgaatctctcttgagaagagatctgataacacacacacacacacacacacacacac 2220
|||||
DB 2161 GACGAATCTCTCTTGAAGAGATTTCACTACAAACACACACACACACACACACACACAC 2220
OY 2221 caaccacagctacagctacagctacaggaagcaaaagcacaagctcccaaccacagca 2280
|||||
```

```
DB 2221 CAACCAACGATACAGTACCAAGTACCAAGCAAAAGCACAAAGTCCCAACACACACAGCA 2280
OY 2281 ccagtaataataaactgaaatgctccaaatgagatatacttgaaaaatataatga 2340
|||||
DB 2281 CCAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2340
OY 2341 ttttaataactcattatataatgacaaatataatcttggtttcacacacacacacacac 2400
|||||
DB 2341 TTTTAAATCTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2400
OY 2401 gaaaagatlaaaacacataaaatatacaaaaggaaggaagcaaatlaagttcaatg 2460
|||||
DB 2401 GAAAAATTAATAACCAATTAATAATTAACAAAGAGAGAAAGCAAAATTAAGTTCATGT 2460
OY 2461 gatccatagacttatgcttlaataatacaaaataacacacacacacacacacacacac 2520
|||||
DB 2461 GATCCATTAAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2520
OY 2521 gataagcttaaacatagtttatcaacacacacacacacacacacacacacacacacac 2580
|||||
DB 2521 GATAGCTTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2580
OY 2581 tgaattataataaacttaaggaataatgacacaaatlaaaaaattatagaaggaagcga 2640
|||||
DB 2581 TGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2640
OY 2641 aaagatccacacacacacacacacacacacacacacacacacacacacacacacac 2700
|||||
DB 2641 AAAGTATCAACATCTGTAAAAACCTTTCAAGTCAATTAATTAATTAATTAATTAATTA 2700
OY 2701 cctcagaataaacacacacacacacacacacacacacacacacacacacacacacac 2760
|||||
DB 2701 CCTCAGATTAACCCGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 2760
OY 2761 aatagtttaaaatattctgaaaaacatactgagctctgaaaaaaacacacacacacac 2820
|||||
DB 2761 AATAGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2820
OY 2821 gaattaaatagtcacaaagtaagtgaaactcttaataaagaatataaagaatagtgat 2880
|||||
DB 2821 GAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2880
OY 2881 acattctataatgacacacacacacacacacacacacacacacacacacacacacac 2940
|||||
DB 2881 ACATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2940
OY 2941 ctgaatgataatcaaaaggaagaatagaagaatataaataaataaataaataaataa 3000
|||||
DB 2941 TTGAATGATGAATCAAAAGCAAGAAATTTAGAAAGATTAATTAATTAATTAATTAATTA 3000
OY 3001 ttacagctacatcttggtttcataataataataataataataataataataataata 3060
|||||
DB 3001 TTACAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3060
OY 3061 aaaaagaaacagcttgtaataataaataataataataataataataataataataata 3120
|||||
DB 3061 AAAAAGAAACAGTTGGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3120
OY 3121 caattagaatcaaaattgaattcaattaaataaaccacacacacacacacacacacac 3180
|||||
DB 3121 CAATTAGAATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3180
OY 3181 gtttctcttaacacacacacacacacacacacacacacacacacacacacacacac 3240
|||||
DB 3181 GTTTCTCTTTTAAACAAAAAGAGCTGAATTAATTAATTAATTAATTAATTAATTAATTA 3240
OY 3241 acaaaatattattgaaacattataaagacatttaataataatgagtgatcatct 3300
|||||
DB 3241 ACAAAATATTATTGAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3300
OY 3301 ccaatlaaaactcttaagtgaaagatcaatctcaaaacagaagataatgacagcttgaa 3360
|||||
```

Db 3301 CCATTAAAACTTAAAGTGAAGATCATTCAAACAGAAAGATTAATTGCGCATTAGAA 3360
 Qy 3361 aactttaaagtatttaagtaattagaagaataaaggtaatttaattagaagaag 3420
 Db 3361 AACTTTAAAGTATTAAAGTAAATTGAAGAAATTTAAAGATTAATTAAATTGAGAAAG 3420
 Qy 3421 aaaaatcatcactatcgaagtgagatcacatcatttaattgctgaatlaagaagta 3480
 Db 3421 AAAAAATTCATCACTTATCAACAGTGAATGATCATCTTAATTCCTGATTAAGAAAGTA 3480
 Qy 3481 ataaaaataaaaaataatcatcaggttaattcccaagtgaaataataatcgatgttaaat 3540
 Db 3481 ATAAAAAATTAATAAATTAATACAGGTAATTCCTCAAGTGAATAATACGGTGTGTAACAAT 3540
 Qy 3541 gcaatagaactctacaaaaaatctccccaagaagacagatgttcaacagttgtaagt 3600
 Db 3541 GCATTGAATCTTACAAAAAATTTCTCCAGAAAGAACAGATTTTCCACAGTTGTAAGT 3600
 Qy 3601 gaaagtgcgtccgaacatctagaacaagaatcacaagaagaacagcatcaactcgtga 3660
 Db 3601 GAAAGTGCATCCGACACATTAGAACAAAGTCAACCAAGAACACAGCATCACTCATGTA 3660
 Qy 3661 ggagcgaagctcaacaacatacaacatacaacaagtgcgtgaatgagtagtcagta 3720
 Db 3661 GGAGCAGAGTCTACACAAATACACATCAACAAATGTCGATGATGAAGTAGTACGTA 3720
 Qy 3721 atcatagtaacctataatttggaagaatccgaagaagaatcatcagatctagaagaagta 3780
 Db 3721 ATCATAGTACCTATATATTTGGAGATCCGAAGAAGATTATGATCATTTAGSACAGTAGTA 3780
 Qy 3781 acaagagaagcagtaactccctccglaattgataacatacttcttaaaatgaaatgaa 3840
 Db 3781 ACAGAGAGAGCGAGTAACTCTCCGTAATGATTAACATCTTCTTAAATGTAATGAAATGAA 3840
 Qy 3841 tatggagtttatatttaaaaccttaagaagtggtttatagaagtttaaaaaacata 3900
 Db 3841 TATGAGTATTATTTATTTAAACCTTTAGCAGGTTTATAGAGTTTAAAAAAACATTA 3900
 Qy 3901 gaaaaaacaagctatgacatttaattgtaattgaagatctttaaactcacagattta 3960
 Db 3901 GAAATTAACGTTATGACATTTATGTTATGTTAAGGATTTTAAATTCAGATTTAAT 3960
 Qy 3961 aaagctgaaaaattcaaaaatggtttatagaatcagatttaattccatataagaattca 4020
 Db 3961 AAACGTGAAAAATTTCAAAAATGTTTAGAATCAGATTTAATTCATATTAAGATTTAACA 4020
 Qy 4021 tcaagtaattatggttcaagaatcatataaattcttaataagaagaagaagaataa 4080
 Db 4021 TCAAGTAAATTAATGTTGTCAAAAGATTCATTAATTTCTTAATTAAGAAAAAGAGATTA 4080
 Qy 4081 ttcttaagcagttataatcatataatagaatcagatcagatcagatcagatcagatc 4140
 Db 4081 TTCTTAAGCAGTTATTAATTAATTAATTAAGGATTCATTAAGATTAAGATTTTCCAAAT 4140
 Qy 4141 gatgttcttgatatttaaaaaatcatatcgcgaaaaaataatcagatttagatcaat 4200
 Db 4141 GATGTTCTTGATATTATAAATAATTAATTCGGAATAATTAATTAATTAATTAATTAAT 4200
 Qy 4201 aaaaatcatcagaacagaagaagaagaagaagaagaagaagaagaagaagaagaaga 4260
 Db 4201 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4260
 Qy 4261 attgagacctataaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 4320
 Db 4261 ATTGAGACCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4320
 Qy 4321 gcaaaagtctcaaatcatatcagaaatcgaacgtaagaagtaaaataaagaagact 4380
 Db 4321 GCAAAAGTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4380
 Qy 4381 aattacttaaaaaaatcaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 4440
 Db 4381 AATTACTTAATAAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4440

Qy 4441 ggaatgctgattatcaacaagaattataacataaataactatgacaaagtccctagt 4500
 Db 4441 GGAATGCTGATTATTAACAAGATTAATTAACATTAATTAATTAATTAATTAATTAATTA 4500
 Qy 4501 acagatagttcttgaaatctgctgaagaacgctttatcctaattacttgaatgaac 4560
 Db 4501 ACAGATAGTCTTTTGAATAATCTTGCTAAACCGTTTATCTTAATTAATTAATTAATTA 4560
 Qy 4561 tgcgaagatglttaaacatctcaacaacccaatgcgtlaaaaaaacaatgctcacaaat 4620
 Db 4561 TTCGAAGGATGTTTAAACAATTTCCACAACCAATGCGTAAAAAAACAATGTCACAAAT 4620
 Qy 4621 tctgagttcttgagacatttagtgaagaagaagaatgtaaaatglttatataatcaaa 4680
 Db 4621 TCTGATGTTTCCAGCATTTAGATGAAGAAGATGTAATGTTTAAATTAATTAATTAATTA 4680
 Qy 4681 caagaagtgatataatglttgaataatccaatccctactgtgaagaagaagaatgltga 4740
 Db 4681 CAAGAAGTGATTAATGTTGTAATAATCCAAATCCTACTTGTACCAAAATAATGTTGGA 4740
 Qy 4741 tgtgatgcagatgccaatgltacagaagaagaatcagtagcaacggaagaaatcacaa 4800
 Db 4741 TGTGATGCAGATGCCAAATGTAACCGAAGAAGATTCAGTAGCAACGGAAGAAATCACA 4800
 Qy 4801 tctgagttcttgaaacactgattctatccacttttcgattgatttctgcagttccct 4860
 Db 4801 TGTGATGTTCTTAACCTGATTTTATCCACTTTTCGATGATGTTTTCGAGTTCCTCT 4860
 Qy 4861 aacttctagaatcatcattcttaatacactcatglttaataatlaatacattcatlaa 4920
 Db 4861 AACTTCTGGAATATCATCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4920

RESULT 2

PEANT195 5917 bp mRNA INV 30-MAR-1995
 LOCUS
 DEFINITION Plasmodium falciparum mRNA for major surface antigens precursor (P195).

ACCESSION X02819.1 GI:9864
 VERSION
 KEYWORDS antigen: direct repeat; signal peptide; surface antigen.
 SOURCE malaria parasite P. falciparum.
 ORGANISM Plasmodium falciparum

REFERENCE 1 (bases 1 to 5917)
 AUTHORS Holder, A.A., Lockyer, M.J., Odink, K.G., Sandhu, J.S., Riveros-Moreno, V., Nicholls, S.C., Hillman, Y., Davey, L.S., Tizard, M.L.V., Schwarz, R.T. and Freeman, R.R.
 Primary structure of the precursor to the three major surface antigens of Plasmodium falciparum merozoites.

TITLE
 JOURNAL Nature 317 (6034) :270-273. (1985).
 MEDLINE 8601435
 REFERENCE 2 (bases 1 to 5917)
 AUTHORS Holder, A.A.

JOURNAL Direct Submission
 Submitted (06-MAR-1991) Holder A.A., National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK
 REMARK Update of published sequence
 COMMENT Data kindly reviewed (06-MAR-1986) by R.R. Freeman.

FEATURES
 Location/Qualifiers
 1..5917

sig_peptide
 418..474
 /product="pul. signal peptide (aa -19 to -1)"
 418..5337
 /note="P195 precursor"

CDS
 /codon_start=1
 /protein_id="CAA26676.1"
 /db_xref="GI:9865"
 /db_xref="SWISS-PROT:P04933"
 /translation="MKTIIFLCSFLPIINTGCVTHESYQELVKKLEALDENVLTGTS"


```
|||||
Db 1618 AATGAATATCCAAATGAGAGTACTTATCCCTTTCATATACGATATTTAAACAATGCTTTA 1677
QY 1261 aatgaactaatcttcttggltgaatlaaatccatltgatatacaagaaccagc 1320
Db 1678 AATGAACCTTAATCTTGTGGATTTAATTAATCCATTTGATTAACAAAGAACCAAGT 1737
QY 1321 aaaaacatatactgataatgaagaanaaaatcctaataatgaatlaaggaanaa 1380
Db 1738 AAAACATATATCTGATATGAGAGAAAAATTCATTAATGAATTAAGAAAAATTT 1797
QY 1381 aaaaatgaanaaaaaaaaltgaatcgtataaaaaatcctaagaagacagatcga 1440
Db 1798 AAAATGAAAAAATAAATTTGAATCTGATTAATAAATCTTACGAGACCAATCTAAGTCT 1857
QY 1441 ttaaatgataaacaagaatgaagaatlaagaatlaagaatlaagaatlaagaat 1500
Db 1858 TTAAATGATTAACAAAGATATGAAAAATTTACTTAATGAATTAATGATGACCAATTC 1917
QY 1501 aataataatagatcttaacttaattcgaanaaatgaatggtaaaagataatcataa 1560
Db 1918 AATAAATATATGATTTAACTTAATTTGAAAAAATGATGGTAAAGATTTTCATATTA 1977
QY 1561 gttgagaacttaacacacaaatlaactlttgcaatcctaagaatcttaacaaat 1620
Db 1978 GTTGACAAACTTACACACCATTAATACTTTTGATCCATGAAAAATCTAATAACATTAATCT 2037
QY 1621 gaaaagttaacaaaagctcttaatatagagaatattatcttlaaggaatagtagt 1680
Db 2038 GAAAAATTAACAAAGCTCTTAATATATGGAAGATTAATCTTTAAGGAATTAATGATGTT 2097
QY 1681 gaaaagaatlaaataatlaaanaatlaataagcaaaaatagaanaatgaatga 1740
Db 2098 GAAAAAGATTTAAATATTTAATAAATTTTAATAGCAAAATGACAAATGATGTAAGCA 2157
QY 1741 ttaagtgaataatcttaaaaaagatgaagacagcttctgaaaaaaaatlaactaa 1800
Db 2158 TTATGTTGAATTAATTTAAAAAGATGAAAGAACCTTTTTCAAAAAATAATTAATAAGAC 2217
QY 1801 gaaaataaaccaagatgaanaaatltgaagaatctgacagatltgaagaatga 1860
Db 2218 GAAAAATTAACCGATGAAAAAATTTTGAAGATCTGACATTTGTAAGTAAGTACAA 2277
QY 1861 aaagtcttaataagacaanaatltgaacgaatlaaaaaagatcgaatgtatctta 1920
Db 2278 AAAGTTTATTAATGAACAAATTTGACGAATTAATAAAGCTCAATTTGATTTAAAAAAT 2337
QY 1921 gtagaatlaaacaataataatgttcccaatcttcaacaagaanaataagcaaga 1980
Db 2338 GTTAGAATTTAAACATATATATACATGTTCCCAATCTTACAAACAGAAATAATAGCAAGAA 2397
QY 1981 ccttatatttaattgtgtgaaaaaagaatgtatgaatlaaagtgttcatgacct 2040
Db 2398 CCTTATTTATTTAATTTGTGTTGAAAAAAGAAATTTGATTAATTTAAAGTGTTCATGCCCTAAG 2457
QY 2041 gtagaatctatgaataatgaagaanaaaaaacataaaaaacagaagtctcaatg 2100
Db 2458 GTTAGAATTTCAATGATTAATGAGAAAAAATAAATCAATTAAGAGTCAATGATTAAT 2517
QY 2101 tggagaacctcaacgaaggaagaaataacagagcaagcaactcaaaactgtgaca 2160
Db 2518 TCGGAACCATCAACCGAAGGAAATTAACGAGCAAGCAACTCAAAACCTTGAGCAACAA 2577
QY 2161 gcaagatctgctttagaagagagatcagtaacagcaaacacagaagaaacaaagca 2220
Db 2578 GCAGAGTCTGCTTTAGAGAGAGATTTAGTACAGCACACACAGAGCAAGCAAGCAAGCA 2637
QY 2221 caaccaacagtlaccagtlaccagtlaccagtlaccagtlaccagtlaccagtlaccag 2280
Db 2638 CAACGACGAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 2697
QY 2281 ccagtaataataaaccgaaaaatgttcccaatagatcttgaataatataagaa 2340
|||||
```

```
Db 2698 CCAATAATTAATAAACTGAAAAATGTTTCCAAATTAAGATTAATCTGAAAAATTAATAGAA 2757
QY 2341 ttttaataactcatatataatgcacaataatlttgglttcaacatcaactagaac 2400
Db 2758 TTTTAAATTAATCTTCATATATATATGTCACAAATTAATTTTGTTCACCTCAGCTATGAAAC 2817
QY 2401 gaaaagatlaaacaataataaatacaagaaggaagaagaacaaatlaagtcaat 2460
Db 2818 GAAAGATTAATAAACATTAATAATTAATAACGAGAGAAAGCAAAATTAATTAATCATGT 2877
QY 2461 gatccatlaagatcttcttgaatlaatacaaaataacatcctgtlaatgtatc 2520
Db 2878 GATCCATTAACCTTAATTTGTTAATTAACAAATAATTAATTAATTAATTAATTAATTAAT 2937
QY 2521 gatagcttaacaaatgattatcacaaacttaataagaaatlaagaagaagaat 2580
Db 2938 GATAGCTTAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2997
QY 2581 tgaatlaataaactlaagataatgaacaaatlaaanaatlaaataatlaagaag 2640
Db 2998 TGTAATTTAATAACTTAAGATTAATGACAAATTAATAAATTTAATTAAGAGAGCAAGAA 3057
QY 2641 aaagtatcaacatctgtlaaanaactcttcaagltcaatcaatgaacatlaata 2700
Db 3058 AAAGTATCCACATCTGTAAAAAATCTTTCAAGTTCATCAATGCAACCATTAATTAATTAAC 3117
QY 2701 cctcagatataacccgaagatgaagaatgaatgaatgaatgaatgaatgaatga 2760
Db 3118 CTTGAGATTAACCCGAGATTAAGTGAATGATGATGATGATGATGATGATGATGATGATGAT 3177
QY 2761 aatagltlaaataatltgaanaacataltgaatccttgaagaanaaaacaaatata 2820
Db 3178 AATAGTTTAATAATTTATTTGAAAAATTAATGAGCTTGAAAAAATAATTAATTAATTAACAA 3237
QY 2821 gaatlaagatgaagaagaatgaagaactcttgaagaagaatlaagaagaatga 2880
Db 3238 GAATTAATTAAGTCAAAAAAGTGAATAAACTTTATGAAAAAGATTAATTAATTAATTAAT 3297
QY 2881 acatlaataatgaatccttlaacaatlttgaagaatlaagaactgaatgaatga 2940
Db 3298 ACATTTTATTAATTAATCTTTTACAAATTTTGTAAATCTTAACCTGATGATTAATTAATTA 3357
QY 2941 ttagatgaatgaacaaagaagaagaatlaagaagaatlaaataatlaaanaa 3000
Db 3358 TTGAATGATTAATTAATAAAGAGAAATTAAGAGAGATTAATTAATTAATTAATTAATTAAT 3417
QY 3001 ttagatgaatgaacaaagaagaagaatlaagaagaatlaaataatlaaanaa 3060
Db 3418 TTACAGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3477
QY 3061 aaaaagaagaagctgtgataatlaaagaagaatlaaanaaactcttataaaga 3120
Db 3478 AAAAAAGAAACAGTGTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3537
QY 3121 caatlaagaatcaaatltgaatcacttaataacaaagaatltatlaaanaact 3180
Db 3538 CAATTAAGATTAATAATTTGAATTTCACTTAATTAACCAAGATGATTAATTAACCAAACTTTCT 3597
QY 3181 gtttcttlaaacaagaagaagaatgaagaatgaagaactgaagaacacatlaaga 3240
Db 3598 GTTTTCTTTAATAAATAAATAAAGAGCTGAATTAATTAATTAATTAATTAATTAATTAAT 3657
QY 3241 acaaaaatlaataatlaagaacatlaaagaactgtlaaataatlaaataatlaaga 3300
Db 3658 ACATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3717
QY 3301 ccaatlaaanaactlaagtlagaatcaatltcaacaagaatlaataatccagtlaga 3360
Db 3718 CCATTTAAAAAATTAAGTGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3777
QY 3361 aactlaagaatlaagaatlaagaagaagaatlaagaagaatlaagaagaatlaaga 3420
Db 3778 AACTTTAAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3837
|||||
```

OY	4501	acaggtatggttttcttgaaatacctctgcctaacaaccgittacttaattacttcgtatggaac	4560
Db	4918	ACAGGTATGCTTTTGAAATACTTCTTAACCCGTTTTATTCAATTACTTGATGAACAAC	4977
OY	4561	ttagcaagtatgtttaaacatttccacaaccaatcgtaaaaaaacaatgccccaaaat	4620
Db	4978	TTCGAAGGTATGTTAACAATTTCCACAAACCACCATCGTAAAAAAAACAATGTCACCAAAT	5037
OY	4621	tctgatcttccaagacatttatagatgaagaagaatgtaaattgttlaaatcacaa	4680
Db	5038	TCTGATGCTTTTCAGACATTTTAGATGAAGAAGAGAAATGTTATTTTAATTAACAA	5097
OY	4681	caagaaggatgaataatgigtgtgaaaaatccoaactctacttgytaacgaataatgtyga	4740
Db	5098	CAGAAGAGGTGAATAATGTTGGAAAAATCCAATCCTACTTGTGAAGAAAATTAATGCTGA	5157
OY	4741	tgttgtcagatggccaatatgtaccgaagaagaattagtgtagaagaagaaataatcaca	4800
Db	5158	TGTGATGCAGATGCCAAATGTACCGAAGAAGTTTCAGATGACGACGGAAGAAAATCAC	5217
OY	4801	tgtgatctacaaccgcatcttcatcoacttgcgtgattcttcgcagttccct.	4860
Db	5218	TGTGATGATGACAAACCGATTCCTTATATCCACTTTTCGATGATGATTTTCGCACTTCCT	5277
OY	4861	aactcttaggaatactcttactataactcatgtaataatlaaacagttacttaa	4920
Db	5278	AACCTTCTTAGGAATATCAFTTCTATTATFACCAATGTTAATATTATACAGTTTCATTTAA	5337
RESULT	3		
PEGP195A			
LOCUS	7038 bp	mRNA	INV 03-APR-1995
DEFINITION	Plasmodium falciparum mRNA for major merozoite surface antigen gp195		
ACCESSION	X15063		
VERSION	X15063.1	GI:9896	
KEYWORDS	antigen; glycoprotein; gp195 gene; surface antigen.		
SOURCE	malaria parasite P. falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 7038)		
TITLE	Mylers, P.J.		
JOURNAL	Direct Submissions		
REFERENCE	Submitted (14-APR-1989) Myler P.J., Seattle Biomedical Research Institute, 4 Nickerson Street, Seattle WA 98109-1651, U S A		
AUTHORS	2. (bases 1 to 7038)		
TITLE	Mylers, P.J.		
JOURNAL	Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from Plasmodium falciparum Palo Alto PLF-3/B11		
MEDLINE	Nucleic Acids Res. 17 (13), 5401 (1989)		
FEATURES	89345116		
source	Location/Qualifiers		
	1..7038		
	/organism="Plasmodium falciparum"		
	/strain="Palo Alto PLF-3/B11"		
	/db.xref="taxon:5833"		
	/db_stage="trophozoite/schizont."		
	1592..1648		
	/product="signal peptide (AA -19 to -1)"		
	1592..6511		
	/codon_start=1		
	/product="gp195 surface antigen preprotein"		
	/protein_id="CAA33163.1"		
	/db_xref="GI:9897"		
sig_peptide	/translation="MKTIIFPLCSFLPIITMOCVTHESYOELVKKLEDAVLGYRS LFOEKVALINEGSTGTAIVTSTPGSKSVASGGSCVASGGSVAAGSGSVAS GGSSSRRTNDSNDSDAKSYADLEHVRNYLTETKEKYLPOLFDLTNHMLTDGN HGCKRYLDIGYEELINELLYRLNFPEDLRAKLINDVCANDYCOIIPNLKITANELDVYIK KVNGYKKRPDNTKDYNCKMEDYTKRKKTILENINELIESKTTIDKNKAITEEEK NTPAOYDLSITNKOLEAHNLISVLEKRDIOTLKNNITVELDKTHEIRNPANGS KPPTLLDDKKKIEEHKEIKEIAKTIKFNIDSLTFDPLEEYLAREKNKINIDIASV ETKSTENEPNVGVTPLSYNDINNALNELINSFGCLINFPDYTKEPSKNIYDNERK KFINKIEKIKIEKKIESDKSYEDRSKSILNDITKEYEKLNIEYDSKFNNNIDLTN		
CDS			

QY 1741 ttagtltgaaatactaaagaagaagacagcttltgaaaaaaacttaataagac 1800
|||||
Db 3332 TTAGTGAATAATATTAAGATGAAGACACTTTTGAAAAAAATTAATAAGAC 3391
QY 1801 gaaaaaacaagatgaagaatcttaagaagtactgacatgtgaaagtaagttcaa 1860
|||||
Db 3392 GAAAAAACAACAGATGAAAAATTTTGAAGATCTGACATTTGAAAAAGACAGTTCAA 3451
QY 1861 aaagtttlttaatgaacaaatltgacgaatlaaaagaactgaatlttaaaaaat 1920
|||||
Db 3452 AAAGTTTATTAATGAACAAAATTTGACGAATTAATAAGACTCAATTTGATTTAAAAAAT 3511
QY 1921 gtagaattaaacatgatacagtcgtcccaatcttacaacaagaanaaataagcaaga 1980
|||||
Db 3512 GTAGATTTAAACATATATACATGTTCCCAATTTTACAAACAGAAAATAAGCAAGAA 3571
QY 1981 ccttatttlaattgtgtgtgaaaaagaatltgaataatlaaaagtgltcatgtcctaag 2040
|||||
Db 3572 CCTTATTTATTTAATTTGTTGAAAAAGAAATTTGATTAATTAAGTTCATGCTCAAG 3631
QY 2041 gtagaatctgtgataatgaagaagaaaaaaacataaaacagaagtgtaactggaat 2100
|||||
Db 3632 GTAGATTCATTTGTTAATGAAGAAAAAAACATTAATAACAGACGTCATGCTGAAT 3691
QY 2101 tgcgaaccatacaacgaagagaataacacagacgaagcaactacaacctggaacaa 2160
|||||
Db 3692 TCGGAACCATCAACGAGAGAAATTAACAGGCAACGCAATCAAAACCTGACACAA 3751
QY 2161 gcaagatctgtcttagaagagagatltcagtaacagcaacagcaagaacaaacaaagca 2220
|||||
Db 3752 GCAGATCTGCTTTAGAAAGAGATTCAGTAACAAGCACAAAGCACAAAGCAAAAGCA 3811
QY 2221 gaacccacgtaccagtagcagacgaagaagcaagaagtcaccaacacacagca 2280
|||||
Db 3812 CAACCAACGATCCAGTACAGACAGAAAGCAAAAGCAACAGTCCACACACACAGCA 3871
QY 2281 ccagtaataataaactgaataatgttccaaatagatlatcttgaataatataatga 2340
|||||
Db 3872 CCAGTAATAATTAATACTGAATAATGTTCCAAATTTAGATTATTTGAAAAATTAATGAA 3931
QY 2341 tttttaataactcatatataatgcaaaatataatlttggttccacacacacatgaac 2400
|||||
Db 3932 TTTTAAATCTTCATATATCTCAAAATATATTTGTTTCAACATCAATATGAAAC 3991
QY 2401 gaaaataataaacaataaataatcaagaaggaagaagaacgaatgaatcatgt 2460
|||||
Db 3992 GAAATATTTAAACAAATTAATAATTAACAAAGAGAAAGAAAGCAAAATTAATCATGT 4051
QY 2461 gatccatagactatgttlaataatacaaaataacataacctgaatgtatctatgtct 2520
|||||
Db 4052 GATCCATTTAGACTTATGTTTAAATACAAATAATACATACCTGTATGTATTTCTATGTTT 4111
QY 2521 gatagcttaacaatagtttatacaacaactataatgaataatlaagaagaagaatgtct 2580
|||||
Db 4112 GATAGCTTTAAACAAATGTTTATCACAACTATTATGAAATTTTGAAGAAAGAAATGTTT 4171
QY 2581 tattaattataaacttaagaagtaatgaacaaatlaaaatlaataagaagaagcaaa 2640
|||||
Db 4172 TGTAAATTTATTAATACTTAAGATTAATGAACAAATTAATAATTAATTAAGAGAGAGAA 4231
QY 2641 aaagtaatacaactgtlaaaaaactcttcaagtltcaatgcaacatcatatcatcaaa 2700
|||||
Db 4232 AAAGTATTCACATCTGTAAAACTCTTCAAGTTCATCAATGCAACCATTTATCTTAACA 4291
QY 2701 cctcaggaataaacccgaagtaagtgcaaatgataatcatcatcttcaaaatltgaat 2760
|||||
Db 4292 CCTCAGAGTAACCCGAACTAAGTGCAAATGTGATACATTCATCAATTTGAAT 4351
QY 2761 aatagtttaaataatlttaaaaaacataatgaatcttgaagaaaaaagaataatataccaa 2820
|||||
Db 4352 AATAGTTTAAATTTATTTTAAAAACATATTTAGAGTCTTGAAAAAACAATAATATATCCAA 4411
QY 2821 gaataataagtcacaaaagtagtgaanaacttlatgaagaagatatlaaagaatagat 2880

Db 4412 GAATTAATAGCTCAAAAAGTACTGAACCTTTATGAAGATATTTAAAGTACTGAT 4471
QY 2881 acatlttaataatgaatccttltacaatlttgttaaatctaaagctgaatataatca 2940
|||||
Db 4472 ACATTTTATATTAATCTTTTACAAATTTTGTAAAACTTAACCTGATATATTAATCA 4531
QY 2941 ttagatgaatcaaaagaagaatltgaagaagataatlaaataaataaact 3000
|||||
Db 4532 TTGAATGATCAATCAAAAAGAGAAATTTGAAGAGATTTAATTAATAAAAAAAGT 4591
QY 3001 ttaacatcatcatgtatlatataataataataatlaaaatlaagaagatlttgt 3060
|||||
Db 4592 TTACAGTTATCATTTGATTTATATTAATTAATTAATTAATTAATTAATTAATTTGAT 4651
QY 3061 aaaaagaanaacagltgtlaaataataatgtcaaatlaaaaaacttacttataaaga 3120
|||||
Db 4652 AAAAAAGAAACAGTGTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4711
QY 3121 caattgaatcaaatltgaatlttacttaataaccgaagcatgtatcaaaaacttct 3180
|||||
Db 4712 CAATTTGAATCAAAATTTGAATTCCTTAATTAACCCAAAGCATGTATTAACAACTTTTCT 4771
QY 3181 gtttctttaaacaagaaagaagctgaaatagcagaacctgaacacatlagaaac 3240
|||||
Db 4772 GTTTTCTTTTACAAAAAAGAGAGCTGAATATGCAAGAACTGAACACATTTAGAAAC 4831
QY 3241 acaaaatattatgaacattataaaggactgtlaaatalatataatgtgtaacatct 3300
|||||
Db 4832 ACAAAATATTTATTAAGAACCTTTTAAAGAGCTGTTAAATATTAATTAATTAATTAAT 4891
QY 3301 ccatlaaaacttgaatgtgaagaatcaatltcaacgaagaatataatgtccagttgaag 3360
|||||
Db 4892 CCATTTAAACCTTTAAGTGAAGATCAATTTCAACGAAGATTAATTTGCCAATTTAGAA 4951
QY 3361 aacttlaagatlaagtaaatlagagaagaaatlaaagatlaattlaaattlaagaag 3420
|||||
Db 4952 AACTTTAAGTATTAAGTAAATTAAGAGAAATTAAGAGATTAATTTAAATTTAGAAAG 5011
QY 3421 aaaaatatacactatcaaatgaatgtgaatcaatcaatlttaattgtgtgaatlaaaga 3480
|||||
Db 5012 AAAAAATTTATCAATCTTAACAACTGATTAACATCAATTTAATTTCTGAATTTAAAGAG 5071
QY 3481 ataaaaataaataatatacaagtaatltccaaaggaataataatgcgaatgtlaaacat 3540
|||||
Db 5072 ATTAATAATAATAATTTATACAGTAAATTTCTCCAGTGAATAATAATCGGATTTAACAAT 5131
QY 3541 gcatlaaatcttacaanaaatltccccaagaagacagatgtgtcaacagttglaagt 3600
|||||
Db 5132 GCATTTAGATCTTTACAAAAATTTCTCCAGAGAGAAACAGATTTGCCAACAGTGTAAAGT 5191
QY 3601 gaaagtgtatccgaacatataagaagaatltcaacaaagaacagcatcaacta 3660
|||||
Db 5192 GAAGATGTATCCGACACATTTAGAAACAACTCAACCAAGAAACCAACATCACTCATGTA 5251
QY 3661 gtagagagatcttaacaaataaacaatacaaaaatgtatgaatgaatgaatgtcgt 3720
|||||
Db 5252 GGAGCAGAGCTTAACCAATTAACACATCACAAAATGTCGATGATGAAGTACGTCGTA 5311
QY 3721 atcatagtaactatatttgaagatccgaagaagatataatgaatlttaagaagtagta 3780
|||||
Db 5312 ATCATGTATGATCTTAATTTGAGAGATCCGAAGAAAGATTAATGATGATTTAGACACAGTAGTA 5371
QY 3781 acaggaagaacaglaactccttccgtlaattgtataacatacttctaaatltgaanaatga 3840
|||||
Db 5372 ACAGGAGAGAGCTAATCTCTCCGTAAATTTGATTAACATCTTTCTAAATTTGAAATGAA 5431
QY 3841 tatgaagtttataatlttaaaactttagcagagtggttatagaagttlaaaaaacata 3900
|||||
Db 5432 TATGAGTTTATTAATTTAAACCTTTAGCAAGGTTTATAGAAAGTTTAAAAAAACATTA 5491
QY 3901 gaaaaaacagttatgaacttaaatgtlaaatgtlaaagatatlaaactcacagattcaat 3960
|||||

Db 5492 GAAAAATACGTTATGACATTTAAATGTTAAATGTTAAAGGATATTTTAAATCAGCATTTAAAT 5551
 QY 3961 aaacgtgaacattccaaaagtctttagaaccagatttaattccataaagaattaca 4020
 Db 5552 AAACGTGAACATTTCAAAATGTTTAAATGATCAGATTTATTCATATTAAGATTNACA 5611
 QY 4021 tcaagtaataatctgtcaagaatccataaaattcttaataaagaanaagataaa 4080
 Db 5612 TCAGATTAATATGTTGTCGAAGATCCATATAATTTCTTAATAAGAAAAGAGATAAA 5671
 QY 4081 tttttagagcgttataatattttagagattcagattcagatataaatttgcgaat 4140
 Db 5672 TTTCTTAAGCGATTATATTAATTAAGATTCAATACGATTAATAATTTTGGCAAT 5731
 QY 4141 gacttccttgatataataaataataataataataataataataataataataata 4200
 Db 5732 GAGTCTTGATATATTAATAATTTATTCGAAAAATATAAATCAGATTAGATTCAATT 5791
 QY 4201 aaaaaataatacaagaacaagaagtgaagaataataataataataataataata 4260
 Db 5792 AAAAAATATATCAACGACAAACAGGTGAATAATGAGAAATACCTCCCTTTTAAACAAAT 5851
 QY 4261 attgaagacctatataaaagaatgaataataataataataataataataataata 4320
 Db 5852 ATTGAGACCTTATATTAATAAGATTAAATGATTAATGATTAATGATTAATGATTA 5911
 QY 4321 gcaaaagcttcaataataataataataataataataataataataataataata 4380
 Db 5912 GCAAAAGTTTAAATTAATATGATTAATGATTAATGATTAATGATTAATGATTAAT 5971
 QY 4381 aattacttaaaaaaattcaagaacaatttgccagattttaaataataataataata 4440
 Db 5972 AATTACTTAAAAACAATTCAAGCAAAATGCGAGATTTTAAAAAATAACAATTTCGTT 6031
 QY 4441 ggaattgcttgattatcaagaagatttaacataataataataataataataata 4500
 Db 6032 GGAATGCTGATTTATCAACAGATTATTAACCAATTAATTAATTAATTAATTAAT 6091
 QY 4501 acagatagcttcttgaaactctgtcaaaacgcttttataataataataataata 4560
 Db 6092 ACAGATATGTTTGTGAAATCTTGTCTAAACCGTTTATCTAATTTACTTGTGAAAC 6151
 QY 4561 tgcgaagatattttaaacaattccacaacaaatgcgttaaaaaaacaatgtccaca 4620
 Db 6152 TTGCAAGGATGTTTAAACATTTCAACACCAATGCGTAAAAAAACAATGTCCACA 6211
 QY 4621 tctgagatcttccagacatttgaagaagaagaatgttaatttataataataata 4680
 Db 6212 TCTGATGTTTTCAGACATTTTGAATGAAGAAGAAATGTAATGTTTAAATTTACAAA 6271
 QY 4681 caagaagtgataaatctgtctgaataatccaaactccactctgaagcaaaataatgtgga 4740
 Db 6272 CAAGAAGGTATTAATCTGTTGAAATCCAAATCTTAATCTTAACCAAAATTAATG 6331
 QY 4741 tctgagatcagatgcgaatgttacgaagaagaatgcagtagcaacgaagaagaata 4800
 Db 6332 TGTGATGAGATGCCAAATGTACCAAGAAAGATTACAGTGAACGAAGAAAGAAATCACA 6391
 QY 4801 tctgagatcagatgcgaatgttacgaagaagaatgcagtagcaacgaagaagaata 4860
 Db 6392 TGTGATGAGATGCCAAATGTACCAAGAAAGATTACAGTGAACGAAGAAAGAAATCACA 6451
 QY 4861 aactctttagaataataataataataataataataataataataataataataata 4920
 Db 6452 AACTCTTAGAATATCATTTCTTAATTAATCAATGTTAATTAATCAATGTTCAITTTAA 6511

RESULT 4
 LOCUS A04562 5760 bp DNA PAT 13-JUL-1993
 DEFINITION P.falciparum P.195 gene.
 ACCESSION A04562
 VERSION A04562.1 GI:410754

KEYWORDS malaria parasite P. falciparum.
 SOURCE Plasmodium falciparum
 ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 5760)
 AUTHORS Holder, A.A., Sandhu, J.S., Odink, K.G., Lockyer, M.J. and Riveros-Moreno, V.
 TITLE Cloning of DNA for protozoal antigens
 JOURNAL Patent: EP 0154454-A 2 11-SEP-1985.
 THE WELLCOME FOUNDATION LIMITED
 FEATURES
 source location/Qualifiers
 1..5760
 /organism="Plasmodium falciparum"
 /db_xref="taxon:5833"
 216..5180
 /gene="P.195"
 216..5180
 /gene="P.195"
 /codon_start=1
 /protein_id="CAA00367.1"
 /db_xref="GI:410755"
 /translation="MKIIFPLCSPLFIITNOCYTHESVOELVKKLEALEDAVLVGLS
 LPOKEMVINEGTSATVTTSTPGSKGASGSGSVASGSGSVASGSGSVASGSGSVAS
 GSGGNSRRTPNSDSSDAKSYADLKHVRNRYLTITKLPOLDELTHNMLTCDN
 HGFKYLIDYEINELLYKLNFEDLKLAKINDYCANDCOIPNLKIRAMELDVLEK
 KLVFGKRLDNIKNVGMEDYIKKNNKTIENINELIEESKTTIDKNKNAKEEEK
 KLYOAYDLSTYNKOLEEHNHLSVLEKRIIDPLKNENIKELDKINEIKNPPANS
 NPNTLPKNTKEIEHEKEIKELATIKRINISLTPLELEYIAREKKNITDISKV
 ETKESTEPNENGTPLSTNDINALELNSFGDLNFPDYTKPEPSKNITTDNERK
 KFINETKEIKLEKKIESDKSYEDRSLSNDITREKLELNEDYDSFENNIDLTN
 PERKMGKRSYVLEKLTNPNTFASYSNHNLEKLTALVAMEDYSLRNIVVEKLY
 YKNLSIKLENELEKIKDEDEFEKIKRDEKPEKLTESDIYKVOYVLL
 MKRIDEKLTOLILNKVELKHNHIVNSYKQNKQEPYLLIYKKEIDOLKEMFME
 SLINERKNIKTOGSDNSPESTBEITGOATKGCQOAGSLBEDSQAQOROKOA
 OPPVPEVPEAKAQPPTPAAPVNNKTENVSKIDYELKLOPLNTSTICKTILVSHST
 NMEKILKOYRTKEESKLSQDPLDLFNIONNIPVMSFSDSNITVYHNLYWEK
 EIGYVFLMETYERKEMVCNLKYLKDNDEIKNLBEAKVSTVTLTSSSSQPLST
 PODKPEVANDSTSHSTNLNLSKLFENILSGKNKIYQELIGOKSSNFYEKLIKD
 SDTPFNESTFNSTKADINDLSLNESKRRKLEEDINKTKTLOSPDLYNNYKLE
 RLFDKKRYGKRMOKIKTILKEOLESLNLSNPKHYLONFSPFNKKRAEAEET
 EMTLENTKLLHLYGVLVYNGESSPLTLEESIOEDNYSLENFENKLEGLK
 KDNLMLEKKILSYLSRGHLHJAELEKLVTKNNYGNSSVNNTVNNALESYKFLP
 ECTDYATVVESSGSDTLQESOPKRPASTHVAESNTITTSQWDEVDVLIYLFGE
 SEEDYDLQGVYGEAVTTSVIDNLISKEINYEYLYKPLAGVYRSLKOLENNVT
 FNVNKKDILNSRFNKRNFKNVLESDLPYKDLTSNVYKPYFELNKKERKDFLSS
 VNYIKSDIDPDINPANDVIGYKTIISKKSDLSIKYINDKQENKYLPELNNIE
 TIKTVNDKIDILFVHLEKAVLNTYTESNVVKIKELVLTIDKDLADFKNNFV
 GIADISTDYNHNNLLTKPLSTGMVFNELSVLSMLMKLRLAYVKEFTTPRRKRTMT
 QOSGCFRHLDERECKCLNLYKQESGCKVENSNTCPENNNGCDADAKCTEEDSGST
 GKRTICECTKPCDYPVLSMVIYFCSSNFELGISPLILIMILYFSI"
 BASE COUNT 2565 a 630 c 725 g 1840 t
 ORIGIN
 Query Match 97.1%; Score 4779.2; DB 9; Length 5760;
 Best Local Similarity 98.4%; Pred. No. 0;
 Matches 4886; Conservative 0; Mismatches 33; Indels 47; Gaps 4;

Oy	241	tcagttgcttcaggtygctcagttgcttcaggtygctcagttgcttcaggtygctcagtt	300
Db	456	TCAGTTGCTTCAGGTGGCTCAGTTGGCTCAGGTGGCTCAGTTGGCTCAGGTGGCTCAGTT	515
Oy	301	gcttcaggtygcttcaggtyaattcaagagtcacaaaccttcaggtyaattcaagtyatca	360
Db	516	GCTTCAGGTGGCTCAGGTATTCACAGCTACAAATCCTTCAGATTAATTCAAGTATTC	575
Oy	361	gagtgtaaatcttaagctgatttaaaacacagagtcgaaattacttgyttaactacaa	420
Db	576	GATGCTAAATCTTACGCTGATTTAAACACAGAGTACGAAATTAATCTGTAACTATCAA	635
Oy	421	gaatccaatatccccaaccttggttttaactaatcatgttbaacttttgygataat	480
Db	636	GAATCCAAATATCCCTCAACTGTTGATTTAACTAAATCATATGTATTAACGTTGTGATAT	695
Oy	481	atcatagcttccaatacttaatttgaatggatatgaaagaattaaatgaattatataaa	540
Db	696	ATTCATGTTTCAAAATATTTAATTTGATGGATTTGAAGAAATTAATGATTTTATATAA	755
Oy	541	ttaaacttttaatttgatttataatgaagcaaaaataatgatgatagtgtgctaagtatac	600
Db	756	TTAAACTTTAATTTGATTTATTTAAAGCAAAAATTTAAAGATGTATGTCTAAATGATAT	815
Oy	601	tgtaaaataccttcaatcttaaatcttgctgcaatgaaattgaagtaacttaaaaaact	660
Db	816	TGTCAATTAACCTTTCATCTTAAATTCGTGCAAAATGAAATTTAGCTTAAAAAACTT	875
Oy	661	gtgttcgatatagaanaaacatagacaatatlaaagaatagttagaanaaaatgynaagat	720
Db	876	GTGTTCGATATGAAACCCATTGACAAATTTAAAGATTAATGTGGAAGAAATGGAGAT	935
Oy	721	tacattaaaaaaataaaaaaaccttagaanaatatgaattgaatttaattggaagaagtga	780
Db	936	TACATTAAAAAATTAAAAAACCATTAGAAATTTAATGATTTAATTGAGAGANGTAG	995
Oy	781	aaaaaatctgataaanaataaagaatlgcaacttaagaagaagaaaaaataatlaacca	840
Db	996	AAAACAATTGATTAATAATAATAGATGCACTTAAGAAAGAAAAAAAATTTATACAA	1055
Oy	841	gtctaatatgattcttctatctatccaataaacaattagaagaagccataattataagc	900
Db	1056	GCTCAATATATCTTCTTATTTTCAATTAATAACAAATTAAGAAAGCAATATTTATATAGC	1115
Oy	901	gtcttagaanaacgattatggcacctttaaanaaaaatgaaacattaaagtaacttgat	960
Db	1116	GTTTTAGAAAAACGTATTTGACACTTTAAAAAAAATGAAACATTTAAGGAATTTACTTGAT	1175
Oy	961	aagataaactgaatlaaaaaatccccccacgycgaactctgaaatataccaaatacttc	1020
Db	1176	AAGATTAATTAATTAATAAATCCCCCACCGGCCAATTTGCAAAATACCAAAATACTCTC	1235
Oy	1021	cttgaataagaacaaaaaaatcgaaggaacacgaaaaagaataaagaagaattgccaanaact	1080
Db	1236	CTTATATAGAACAAAAATCGAGAACACGAAAAAGAAATTAAGAAATTTGCCAAACT	1295
Oy	1081	attaaatttaatactgatagttttattacttgatgccactgaaattgaaatactactatta	1140
Db	1296	ATTAAATTTAATTTATGTACTTTATTTACTTGATCCACTTGAAATTAAGATTAATTTAAG	1355
Oy	1141	gaaaaaaaataaaaatatltgatatagaatlgcaaaaggttgaaacaagaatcaactgaacc	1200
Db	1356	GAAAAAAATTAATAATGTATTAATAGTGCAAAAGTTGAACAAAGAAATCAACTGAACCC	1415
Oy	1201	aatgaatattccaaatggaagttaacttcccttgcataataagaatlaatacaatgattta	1260
Db	1416	AATGAATATCCAATGAGATTTACTTATCTCTTGTCATATAAGATTAATTAACAATGCTTTA	1475
Oy	1261	aagyaacttaactcttltvgtgatttaattatcoatttgattatatacaaaagaaccaagt	1320
Db	1476	AATGAATTAATCTTTTGGTGATTTAATTAATTCATTTGATTAATGAAAGCAACCAAGT	1535

QY	1321	aaacaatatactgctgtaataatgaagaataaaatctcaataatgaaatcaagaataaaat	1380
Db	1536	AAAAACATATTTACTGTAATATGAAGAAAAAAATTTATTAATGAATTAAGCAAAAAATTT	1595
QY	1381	aaatagaataaaaaaaatctgaatctgataaaaaatcttaagaagacagatctaaagct	1440
Db	1566	AAATATGAAAAAATAATTTGAATCTGATTAATAATCTTACGAAACAGATCTTAAGCT	1655
QY	1441	ttaaatgataaacaaaaagataatgaaaaaatctaactaaatcttaatgataatgataagcattc	1500
Db	1656	TTAAATGATATTAACAAAAGATATGAATAATTTCTTAATGAATTTATCATAGCAAAATTC	1715
QY	1501	aataataatagatcttaactaaatctcgaaaaaataatgagtggttaagaatatacataaa	1560
Db	1716	AATATATATATATGATTTTAAGTAATTTGCAAAAAATGATGGTAAAGATATTCATATATA	1775
QY	1561	gttaggaacttaacacccgaataactctggatccctaaagaatctcaaacataatct	1620
Db	1776	GTGGAACACTTACACACCGTAATACTTTTGCACTTAATGAATAATCTTAACATATATCTT	1835
QY	1621	gaagaatgaacaaagctcttaataatataatgaagatactcttaagaatataagtt	1680
Db	1836	GAAAGTTAAACAAAGCTTTAAATATATGAGAGATTTATCTTTAAAGAAATATTAAGT	1895
QY	1681	gaaaaaagataaaaaatattataaaatcttaataagcaaaatgaanaatgagatgaaca	1740
Db	1896	GAAAAAGAAATTAATTTATTTAAAAATTTATATGACAAATATGAAATATGATGTAAACA	1955
QY	1741	ttagttgaanaatatlaaaaaagatctgaagaacagcttttgaaaaaaaaatcttaagaac	1800
Db	1956	TTAGTTGTAATAATTTAAAAAGATGAAGAACACCTTTTGAAAAAATAATTTACTAAAC	2015
QY	1801	gaaaataaacccagaatgaaaaaatcttagaagatactgcgaatctgataaagtataagttca	1860
Db	2016	GAAATATTAACAGATGAATAAAATTTTAAGAATATCTGACATTTGTAAGTACAAAGTCCA	2075
QY	1861	aaagtattataatgaacaaaatctgaacgaatgaanaagactcoaatgatttlaaaaaat	1920
Db	2076	AAAGTTTTATTAATGAACAAATTTGACGAATTTAAAAAACATCAATTTGATTTAAAAAAT	2135
QY	1921	gtagaatlaaagaataatatactgtcccaatcttcaaaaagaanaatagaagaata	1980
Db	2136	GTAGATTAATAACATTAATATATCATGTTCCAAATCTTCAAAACAAGAAATATAGCAATA	2195
QY	1981	ccttaataatctgtctgtgaaaaaaagaatgataaataaaagtgtctatgctctag	2040
Db	2196	CCTATTTATTTAATTTGTTGTTGAAAAAAGAAATTTGATTAATTAAGTGTTCATGCTTAAG	2255
QY	2041	gtagaatcatgtataaataatgaaaaaaaacataaaaaatgaaagttcaatcgtataat	2100
Db	2256	GTAGATCATTTGATTAATATGAAGAAAAAATAACATTAATAACACAGGTCAATCGGATAT	2315
QY	2101	tcggaacatcaaacccgaagaagaataataacagacaaagcaactatacaaatctggaacaa	2160
Db	2316	TCGGAACCATTAACCGAAGAGAAATTAACGCAAGCAATATCAAAAATCTGGACATCAAA	2375
QY	2161	gaagaatctgtcttagaagaagaatctgaatacaagacaaagacaaagaacaaaaacaaag	2220
Db	2376	GCAGATCTGCTTTAGAAAGAGATCTCATCAAGCAACAGCAACAGCAACAAAAACAACA	2435
QY	2221	caaccacagataccagataccagataccagaagcaaaaacaaagctcccaacacacacaaag	2280
Db	2436	CAACACACAGTACAGTACCAAGTACCAAGCAAGCAAGCAAGTCCCAACACACACACACA	2495
QY	2281	ccagtaataataaaactggaanaatggttccaaatagatatctctgzaaaatatatagaa	2340
Db	2496	CCAGTAATAATAAAACGTGAATAATGTTTCCAAATTTGATTAATCTTGAAAAATTAATATA	2555
QY	2341	ttttaaataactataataatgtacaaaatatatttggtttcaactcaactaatgagac	2400
Db	2556	TTTTTAATTAATCTTATATATATGTCACAAAATATTTTGCTTTCACTCACTCAATATGAC	2615
QY	2401	gaaaagataataaacatatataatctacaagaggaagaagaacaaatlaagttcaatgt	2460

Db	2616	GAAGAAGATTTAAACATATTTAAATTACAAAGAGAGAAAGAAAGCAATTAAGTTCATGT	2675
Oy	2461	gaccatagagctatctgtttaataatacaaaatacaataccctgttaatctatcttgltt	2520
Db	2676	GATCCATTAGACTTATGTTTAAATATATACAAATAATACATACCTGTATGATTCCTATGTTT	2735
Oy	2521	gatagctaaacaaatgtttatacaacatct	2551
Db	2736	GATAGCTTTAAACCTTAAGTTTATATACAACTATATATGTTTATGTTTATATATTAAT	2794
Oy	2552	-----tattggaattataatgaagaagaatggtcttataattataaact	2597
Db	2795	TGTATTTATATATCTTATTGCAAAATTTATGAAAAAGAAATGGTTTGTATTTATTAATCT	2854
Oy	2598	laagataaattagacaaaataaaaaattattatagagagaagcgaanaaagtatccacatcgt	2657
Db	2855	TAAAGATATATGACAAATTTAAATTTATTTAGAGAAAGGAAAAAGATACCATCTGCT	2914
Oy	2658	aaaaactcttcaagctcatcacaatgcaacacatatacaataccctcagataaaaccca	2717
Db	2915	AAAAACCTCTTCAAGTTCATCAATGCAACCATTTATTAACCTCGAGATTAACCCGA	2974
Oy	2718	agtaagtgcaaatatgatcatcacaatctacaaatttgaaataagttaaaattat	2777
Db	2975	AGTAAAGTCAAAATATGATATCATCATCTTCACAAATTTGAAATATATAGTTTAAATTAAT	3034
Oy	2778	tgaaacacatcttgagctcttgaaaaaacaanaatacacagaattatagtgcaaa	2837
Db	3035	TGAAACACTTTTGAGCTGTGGAaaaaaCAAAATATATACAGAAATTAATAGTCAAA	3094
Oy	2838	aagtagtgaaaactttagaaaaagabataaaagatagtgatgatacttataatgaaac	2897
Db	3095	AAGTAGTGCAAACTTTTATGAAAAAGATTTAAAGATAGTATACATTTTATATGAATVC	3154
Oy	2898	ttttaacaatttgglaaaatctaagctgtgatatlaattatgaaatgaaatcaaa	2957
Db	3155	TTTTACAAATTTGTAAATCTAAAGCTGATGATTAATTAATGAAAGATGAATCAAA	3214
Oy	2958	aaggaagaatttagaagaagatatataaatcaaaaaaacttacaatgatacttga	3017
Db	3215	AAGGAGAAATTTACAAACAGATTTAATTAATTAATAAAAACTTTACACTATCAATTGGA	3274
Oy	3018	ttctataataaataataaataatbagaaatatttggataaaagaanaacagtttg	3077
Db	3275	TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3334
Oy	3078	taaatataaaatgcaaaatataaaaacttctataataaagaacaattagaatcaaaat	3137
Db	3335	TAAATATTAATAATTCGAAATTTAAAAACCTTACTTTTAAAGAACAAATTAAGATCAAAATTT	3394
Oy	3138	gaatcaacttaataaccaagaacatgatatcaaaaacttctgttctcttaacaaaaa	3197
Db	3395	GAAATTCACCTTATATACCCCAAGCATGATTTACAAAACCTTCTGTTTCTTTAACAAAA	3454
Oy	3198	aaagaagaactgnaaatiagcagaaactgnaaacacatfatagaanaacacaaataatitgaa	3257
Db	3455	AAAAAGAGCTGAATATAGCACAAACTGAAAAACATTAGAAAAACAAAAATATTTATGAA	3514
Oy	3258	acattataaagagcttgttaataataataatggttgaaatcctccattaaaaaacttaag	3317
Db	3515	ACATTATTAAGAGCTTTTAAATTAATTAATTAATGCTGATCATCTTCATTTAAAAACTTTAAG	3574
Oy	3318	tgaagaatcaactcaaacagagaataatattgccaatttagaagaacttgaagatitgaag	3377
Db	3575	TGAAGATCACTTCAATTAACACAGATATTTATGCTGCACTTTAGAAAACTTTAAGATTTAAG	3634
Oy	3378	taaattagaagaagaattaaagataatctbaaatttagaagaagaanaaataatactact	3437
Db	3635	TAAATTTAGAGAAATTTAAAGATTAATCTTAATTAATTTAGAAAAAGAAAAATTAATCACTTT	3694
Oy	3438	atcaagtgatataccaatlaaatgcgcgaattaaagaatataaaaaataaaatla	3497

Db	3695	ATCAAGAGGTTTACATCATTTTAATTCGTGATTTAAAGAAGATATAAAAAATTA	3754
Oy	3498	tacsgtaattcttccaaagtgaaataataacgagtglttaacaatlgaacttaca	3557
Db	3755	TACAGGTAATCTCCAAAGCTTAATATACGAGTGTAAACAATGCAATTAGATCTTACA	3814
Oy	3558	aaattctcccggaaggaacagatgttgaaacagttgtaagtgaaagtgatccgac	3617
Db	3815	AAAATTTCTCCCAAGAAACAATGTTGCAACAGTTGTAAGTGAAGGATCCGACAC	3874
Oy	3618	atlaagaacaagtlaacaaagaacacagcatcaactcgtltagagcagacltaaac	3677
Db	3875	ATTGACAAACAAGTCAACCAAGAAGAACAGATCACTGATGAGCAGCATCTACAC	3934
Oy	3678	aataacaacatcacaaaaatgcgatgaagtagatgaacgttaactaagtac	3737
Db	3935	AATTAACAACATCACAAATATGCGATGATGAAGTAGACGTAATCATAGTACTATTT	3994
Oy	3738	tggagaatccgaagaagatatgatattaggaacagtagtaacagagaagacgtaac	3797
Db	3995	TGGAGAAATCCCAAAACAATATGATGATTTAGCACAGTAGTAACAGGAACGTAAC	4054
Oy	3798	tctctccgtatgtataacgactcttciaaaattgnaaaatgaaatcttgagtttaatt	3857
Db	4055	TACTTCGTAATGTATACATACCTCTCTTAATAATTGAATATGAAAGTTTATATT	4114
Oy	3858	aaaaccttgaacggtgtttataagaagtttaaaaaacaattagaataaaccgttatgc	3917
Db	4115	AAAACCTTTAGCAGCGTGTATTAGAAAGTTTAAAAAACAAATTAGAAATTAACGTTATGC	4174
Oy	3918	atttaagttaagttaagatatattaaatlccagatttaataacgltgaattcaa	3977
Db	4175	ATTTAATGTTAAATGTTTAAAGGATATTTTAAATTCACATTTAATAACCGTAAATTTCAA	4234
Oy	3978	aaatggtttgaacacagatttaattccataaagaattaaacataagaaatataagtt	4037
Db	4235	AAATGTTTGAATCAATTTAATTCATTAAGAATTTTAACATCAAGTAAATTAATGTTGT	4294
Oy	4038	caagaatccatataaactttcttaataagaataaagaataaattcttaagcagttaa	4097
Db	4295	CAAGATCCCATATTAATTTCTTAATAAGAAAAACAAGATAATTTTAACGAGTTATAA	4354
Oy	4098	ttataatgaagatccaatagatacagatataaaatttgcaaatgaltgtctgatat	4157
Db	4355	TTAATTTAAGGATTCATTAATGATACGATATTAATTTTGCAAATGATGTTCTTGATATTA	4414
Oy	4158	taaaatttatccgaaaaataataacacagattagatttaatttaaaaaatatcaaga	4217
Db	4415	TAAATATTTATCCGAAAAATTAATTAATTCAGATTTGATTAATTTAAATAATTAATACGA	4474
Oy	4218	caaaacagtgaaaaatgaagaataccttccctttttaaacaatcttgagacctataa	4277
Db	4475	CAAAACAAGTGSAATATGAGAAATACCTTCCTCTTTTAAACATATTGAGACCTTATATA	4534
Oy	4278	aacsagttaatgaataaattgatttattctgtaattcttagaagcaaaagtctaaat	4337
Db	4535	AACAGTTAATGATTAATTAATTTGATTTATTGTAATTCATTTAGAAGCAAAAAGTCTTAATTA	4594
Oy	4338	tactatagagaataccaagtagaagttaaaataaaagaacttaattacttaaaaacat	4397
Db	4595	TACTATATGAGAAATCAAAACGTACAGCTTAATAATTAAGAACTTAATTTCTTTAAACAAAT	4654
Oy	4398	lcaagaacaattgacagattttaaaaaataacaatttcgtctggaattgctgattc	4457
Db	4655	TCAAGACAATTTGGCAGATTTTAAAAAAATATACATTTCTTTGGATTTGCTGATTTATC	4714
Oy	4458	aacagatatataacataactaactatctgcaaaagtctcttagtaacgltatgatttga	4517
Db	4715	AACAGATTATTAACCATATTAATTAATTTGACAAAGTTCCTTACTACAGGATAGTTTTGA	4774
Oy	4518	aaatcttgctaaaaacogtlttatacttaattacttga-tgnaaaccttgcaggtatgttaa	4576
Db	4775	AAATCTTCTTAATAATCCGTTTATCTTAATTTACTGTGTTGGAATCTGCAAGGTATGTTAA	4834

QY 4577 acattcacacaacacatgcgtaaaaaaacaatg--ccacacacattctgcatgttcag 4634
|||||
Db 4835 ACATTTCACAAACACCAATCGTAAAAAACATGATGACACACAGCTTCGATGTTTCAG 4894
QY 4635 acatttagatgaagaagaatgtaaatgcttatataataacaaagaagtgataa 4694
|||||
Db 4895 ACATTAGATGAAGAAGAGAAATGTAATGTTATTAATAATACAAACAGAGGTGATA 4954
QY 4695 atggtttgaagaatccaaatccctactctgttaacgaataatggtgagtgatgcaatgc 4754
|||||
Db 4955 ATGTGTTGAATTAACAAATCCCTACTTGTACGAAATAATGCGATGTCATGCAAGATGC 5014
QY 4755 caaatgtacgaagaagaatcagtgatgaacgaagaagaataacatgtaagtactaa 4814
Db 5015 CAAATGTACCGAAGAAAGATTCAGGTGACACGAAAGAAATACATGTGATGTAATA 5074
QY 4815 accgtattctatccacttctgcatgtaattcttcgcaattcccttaacttcttaagaat 4874
|||||
Db 5075 ACCGTATGTATCCACTTGTGATGTAATTTCTGACGTCCTTAACCTCTTAGAAT 5134
QY 4875 atcatcttataataactcatgtaattataatatacagttcatttaa 4920
|||||
Db 5135 ATCATCTTATTATACTCATGTATTATTATACAGTTCAATTTAA 5180

RESULT 5
E00656 5760 bp RNA PAT 29-SEP-1997
LOCUS E00656
DEFINITION cDNA encoding P.195 protein of Plasmodium falciparum.
ACCESSION E00656
VERSION E00656.1 GI:2168935
KEYWORDS JP 1986019490-A/1.
SOURCE JP 1986019490-A/1.
ORGANISM Plasmodium falciparum.
Plasmodium falciparum.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 5760)
REFERENCE Ansonil,A.H.,/ Malkuru,J.R., Jiyasubiaa,S.S., Barenteina,R.M. and Karenu,G.O.
AUTHORS

TITLE DNA CLONING FOR ANTIGEN OF PROTOZOA
JOURNAL Patent: JP 1986019490-A 1 28-JAN-1986;
COMMENT WELLCOME FOUND LTD:THE
OS Plasmodium falciparum
PN JP 1986019490-A/1
PD 28-JAN-1986
PF 22-FEB-1985 JP 1985034280
PR 22-FEB-1984 GB 84 8404692, 26-SEP-1984 GB 84 8424340 PI
ANSONIL AASAA HORUDAA, MAIKURU JIEMUSU ROTSUOKUYA, PI
JIYASUBIAA SHINGU SANDEYU, BARENTEINA RIBEROSU MORENO, PI KAREBU
GERTSUTO ODEINKU
PC C12N15/00,A61K39/015,C07H21/04,C07K13/00,C12N1/20,C12P21/02,
PC (C12N1/20,
PC C12R1:19),(C12P21/02,C12R1:19);
CC strandedness: Single;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
FH key Location/Qualifiers
FH CDS 216..5180
FT sig_peptide /product='P.195 protein'
FT mat_peptide 273..5177
FT Location/Qualifiers
1..5760
/organism='Plasmodium falciparum'
/db_xref='taxon:5833'
BASE COUNT 2565 a 630 c 725 g 1840 t
ORIGIN

Query Match 97.1%; Score 4779.2; DB 10; Length 5760;

Best Local Similarity 98.4%; Pred. No. 0;
Matches 4886; Conservative 0; Mismatches 33; Indels 47; Gaps 4;
QY 1 atgaagatcatatctctttttagtcttcttttttttataataataacaaatgta 60
|||||
Db 216 ATGAAGATCATATCTCTTTTATGTTTCATTTCTTTTATTAATAAATACAAATGTGA 275
QY 61 acacagaagaatcatcaagaactgtccaaaactagaagctttaagaatgcaatgct 120
|||||
Db 276 ACACAGAAAGTTATACAACTTGTCAAAACCTTGAAAGCTTTAGAAAGTCAGATTG 335
QY 121 acaggtatagttatcttcaaaagaagaataatgtaataatgaagaaagaatggaaca 180
Db 336 ACAGGTTATAGTTTATTTCAAAAAGAAATAGTATTAAATGAAGAACAAAGTGAAACA 395
QY 181 gctgttaacaactgaacacctggttcaaaaggttcagttgctcaagtggttcagtgac 240
|||||
Db 396 GCTGTACAACTAGTACACCTGGTTCAAAAGGCTTCAGTTGCTTCAGGTGCTGAGTGC 455
QY 241 tcagttgctcaagtggttcagttgcttcaagtggttcagttgcttcagttgctcaagtt 300
|||||
Db 456 TCAGTTGCTTCAGGTGCTTCAGTTGCTTCAGTTGCTTCAGTTGCTTCAGTTGCTTCAGTT 515
QY 301 gcttcaggtggttcaggttaattcaagaagtcacaaatccctcagataatcagttgta 360
Db 516 GCTTCAGGTGTTTCAGGTATTCACAGCAGTACAAATCCTTCAGATTAATTCAGTATTC 575
QY 361 gatgcataatctgaagctgaattaaacacagaagtcgaataactctgttaactcaaa 420
Db 576 GATGCTAAATCTTACCTCATTTTAAACAGAGTACGAAATTAATCTTAACTATCAAA 635
QY 421 gaactcaaatatccatcaactcttgaattcaactacatgtaactctgtgtgtaact 480
|||||
Db 636 GAACTCAAAATCTCTCAACTCTTGTATTAACATTAATGTTAATCTTGTGTATTAAT 695
QY 481 atccatggttccaatatttaattgaatgtaataatgaagaatgaatgaatataataa 540
Db 696 ATTCATGCTTTCAAAATTAATTAATGATGATNGAAGAAATTAATTAATTAATTAATA 755
QY 541 taaactttatcttgatattataagaagcaaaatgaatgaatgtaatgtaactat 600
Db 756 TTAACCTTTATTTTATTTATTTAATTAAGAGCAAAATTAATGAATGATGATGATTA 815
QY 601 tctcaaatcccttcaatctttaaattctgtgaagaatgaatgaatgaatgaatgaat 660
Db 816 TGTCAATATCTTCAATCTTAAATTCGTCAAAATGAATTAAGCTACTTAAATACTT 875
QY 661 ggtctggaatagaagaacatagaacatataatgaatgaatgaatgaatgaatgaat 720
Db 876 GTGTGGAATATGAAGAACCATTAGACATTAATTAAGATTAATAGCAAAATGGAAGAT 935
QY 721 tacatcaaaaaataaaaaaacatagaaatataatgaatgaatgaatgaatgaatgaat 780
Db 936 TACATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 995
QY 781 aaaaacttgataaaaaataagaatgtaactaaagaagaagaagaagaagaagaagaaga 840
Db 996 AAAACATTTGATTAATAATTAAGATGCACTTAAGAAAGAAAGAAAGAAAGAAAGAAAG 1055
QY 841 gctcaatgatcttcttcttcttcaataaacaatgaagaagcacataatgaatgaag 900
Db 1056 GCGCATATATGATCTTCTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 1115
QY 901 gctttagaaaaacgtatgacactttaaataaaaaatgaataatgaatgaatgaatgaat 960
Db 1116 GTTTTGAATAAAGCTATGTACACTTTAAAAAATAAATAAATAAATAAATAAATAAATAA 1175
QY 961 aagataatgaataatgaataatcccaacggcgaatcttggaataatgacacaaatctcc 1020
Db 1176 AAGATTAATGAATTAATAATCCCAACCGGCCCAATTTGGAATTAACCAATTAATCTTC 1235
QY 1021 ctgataagaacaaaaatcgagagacagaagaagaataaagaataatcccaaat 1080
|||||

Db	1236	CTGATAGAACCAAAAAATTCGAGGACACGAAAAAGAAATTAAGAAATTCGCCAAACT	1239
Oy	1081	atcaaatcaatcaltgtaagttattactgataccactggaatagaactatttaaga	1140
Db	1296	ATTAAATTTAATATGTGACTTTATTGTAGTCCACTGGAATPGAAATCTATTAGA	1355
Oy	1141	gaaaaaataaaaatctgtgataagtgcaaggttgaaacaaagaaatccaactgaaacc	1200
Db	1356	GAAAAAAATTAATAATGTGATATPAGTCGAAGGTTGAAACAAAGAAATCACTGAACC	1415
Oy	1201	aatgaataccaagtgaagtaacttaaccttccatacaagacattaaagtctta	1260
Db	1416	AATGATATTCGAATGAGGACTTATTCCTTGTCAATATACGATATTTACATCTTTA	1475
Oy	1261	aagcaacttaactcttctgtgtatctaattaaacatctgattatacaaaagaaacagt	1320
Db	1476	AATGAACCTTAATCTTTGTGGTGAATTAATTAATCCATTTGATTATACAAAGAACCAAGT	1535
Oy	1321	aaaaacatatactcgtgataagaagaaaaaatctcataatgaaatlaagaaaaaat	1360
Db	1536	AAAAACATATPACTGTGATATGGAAGAAAAATTCATTAATGAATTPAGCAAAAAAT	1595
Oy	1381	aaaaagaaaaaaaanaacttgaactcgtataaaaacttcaagagacagactcgaagct	1440
Db	1596	AAATATGAAAAAANAATTTGAACTCGATTAANAATCTTACGAGACAGACTTACGTCT	1655
Oy	1441	ttaaagatatacaaaaagaataagaanaaattacttaagaaattatgatagcaaatc	1500
Db	1656	TTAATATGATATPACAAAGAAAGATATGAAAAATTACTTAATGAATTTATGATAGCAAAATTC	1715
Oy	1501	aataataatagatttaactaatcttcgaaaaaaygaatgggtaaaagatatcatataa	1560
Db	1716	AATATATATATPAGTTTAACTTAATTTTCGAAAAATATGCTAAAGATATTCATATTA	1775
Oy	1561	gttgaagaactgaacacacaaataactcttgaactccatgaanaattcacaacataactt	1620
Db	1776	GTGAGAACTTACACACCTTAATTAATCTTTGCATCTGTGAANAATTTCTTAACATATCTT	1835
Oy	1621	gaaaagttacaagaagctcttaaatatataatgaagaattactttaaagaaatagtaatt	1680
Db	1836	GAAAGCTTAAACAAAAGCTCTTAATATPATGGAAGATTAATCTTTPAAGCAATATAGTACTT	1895
Oy	1661	gaaaaaaatataaataatttaaaaaatttaagaacaaatagaanaatgaaatggaata	1740
Db	1896	GAAAGCAATATTAATAATTTTAAAAATTTTAATPAGCAAAATPAGAAATGCAATTCGAACA	1955
Oy	1741	ttagtgaataatactaaagaatgtagaacagctctttgaaaaaaaattactaaagac	1800
Db	1956	TTATATTGAATATTTTAAAAAAGATGAGACACGCTTTTGAANAANAATTTACTTAAGAC	2015
Oy	1801	gaaaataacccagatggaanaaatttagaagatctgacatgtaaaagtacaagtcca	1860
Db	2016	GAAATATPAAACCATGCAAAAAATTTTAGAAGTATCTGACATGTGAAGTGAAGTTCA	2075
Oy	1861	aaagtcttatataatgaacaaaattgacgaaattaaaaaagactccaattgattttaaanaat	1920
Db	2076	AAAGTTTATTAATGAAACAAAATTTGACGAATPAAAAAAGCTCAATTTATTTAAAAAAT	2135
Oy	1921	gtagaatctaaacataatactacatgtctcccaattcttcaacaaagaanaatatgaagaa	1980
Db	2136	GTAAGATTAATAACATATATATACATGTTCCAAATCTTTCANAAACAGAAATPAGCAAGAA	2195
Oy	1981	ccctattattcaatctgcttgtaaaaaagaanaattgataaattaaagtctcaatgctcaag	2040
Db	2256	CGTTATTTATTTAATTTGGTTGTAAGAAAAAGAAATTTGATTAATTAAGTGTTCATGCTTAC	2255
Oy	2041	gtagaatcattgtaaaatgagaanaaaaaacataaaaaacagaaggtcnaatcggataat	2100
Db	2256	GTAAGATCATTTGATTAATGCAAAAAAANAACATTAATAACCAAGTCATTCGATAT	2315
Oy	2101	tcggaacacctcaaccgagagagaataataacagagacaagcaactacaaaactcgacaaca	2160
Db	2316	TCGGAACCATATCAACGAGAGGAATAATPACAGACACACTTCAAAAACCTCGGCAACAA	2375

OY	2151	gcagagatcctcttagaagagagctcagatcaccaagccaaagccaaagcaaaacaaagca	2220
Db	2376	cgagagatctgccttttagaagagagatcagatcaccaagccaaagccaaacaaacaaagca	2435
OY	2221	caaccaccagatcacagatcaccaagatcaccaagaagcaaaagccaaagctcccaaccacacagca	2280
Db	2436	caaccaccagatcacagatcaccaagatcaccaagaagcaaaagccaaagctcccaaccacacagca	2495
OY	2281	ccagataaataataaacttgaaatgtctccaaatagatctatcttgaaaaatataagaa	2340
Db	2496	ccagataaataataaacttgaaatgtctccaaatagatctatcttgaaaaatataagaa	2555
OY	2341	ttcttaaaacttcataataatctgcaaaatatatttggtctcaactcaacatgagac	2400
Db	2556	tttttttaattaccttcataataatctgcaaaatatatttggtctcaactcaacatgagac	2615
OY	2401	gaaagatataaacaataataaaattacaagaaggaagaaagcaaaatgaattcatgt	2460
Db	2616	gaaagatataaacaataataaaattacaagaaggaagaaagcaaaatgaattcatgt	2675
OY	2461	gacctatgaactatattgtttaatacagaataaataccatgaatattctcatgtt	2520
Db	2676	gacctatgaactatattgtttaatacagaataaataccatgaatattctcatgtt	2735
OY	2521	gatacttaaaacaatagtttatcacaaactt-----	2551
Db	2736	gatacttaaaacaatagtttatcacaaactt-----	2794
OY	2552	-----ttaigaaaattataaagaagaatagtttgtaattataaact	2557
Db	2795	tgctatttaattatcctttatggaataatttataaaaagaaatagtttgtaatttatataaact	2854
OY	2598	taagataatagcaaaataataaaattatctagaggagcgaaaagaatcacacatgt	2657
Db	2855	taagataatagcaaaataataaaattatctagaggagcgaaaagaatcacacatgt	2914
OY	2658	aaaaacctctcaagttcatcaatgcacaacatcatcaatacaacctcagaataaaccga	2717
Db	2915	aaaaacctctcaagttcatcaatgcacaacatcatcaatacaacctcagaataaaccga	2974
OY	2718	agtaagtgcaaatgagatcacatcacatctcaaaatttgaaataatgtttaaaattat	2777
Db	2975	agtaagtgcaaatgagatcacatcacatctcaaaatttgaaataatgtttaaaattat	3034
OY	2778	tgaacaacatctgagctctggaaaaaaacaataatacaagaataataatgagcaaa	2837
Db	3035	tgaacaacatctgagctctggaaaaaaacaataatacaagaataataatgagcaaa	3094
OY	2838	aagtagtgaaaactttatagaaaagatataaagaatgagatataactttataatgatac	2897
Db	3095	aagtagtgaaaactttatagaaaagatataaagaatgagatataactttataatgatac	3154
OY	2898	ttttacaactttgttaaaactctaaagctgataataatcaatcgaaatgataatcaaa	2957
Db	3155	ttttacaactttgttaaaactctaaagctgataataatcaatcgaaatgataatcaaa	3214
OY	2958	aaggagaagaattgaaagaagataataataatlaaaaaaaacttacagttatcaattga	3017
Db	3215	aaggagaagaattgaaagaagataataataatlaaaaaaaacttacagttatcaattga	3274
OY	3018	tttataataataataataataattagaagaatctattgaataaaagaagaacagttg	3077
Db	3275	tttataataataataataataattagaagaatctattgaataaaagaagaacagttg	3334
OY	3078	taaaataaagaatgcaaatlaaaaaacttcttataaagaagaacatagaatccaat	3137
Db	3335	taaaataaagaatgcaaatlaaaaaacttcttataaagaagaacatagaatccaat	3394
OY	3138	gaattcaactataaaccacaagaatgatatcaaaaacttccgttctctttaacaanaa	3197
Db	3395	gaattcaactataaaccacaagaatgatatcaaaaacttccgttctctttaacaanaa	3454

3198 aaagaagctgaatagcagaacggaacacacatlagaaaacacaaaataltatgaa 3357
|||||
3455 AAAAGAGCTGAAATAGCAGAAACGTAACACATTTAGAAAAACAAAAATATTATGAA 3514
3258 aaattataaagactgtttaatatataatgtgtgaatcaatcccatlaaaacttaag 3317
|||||
3515 ACATTATTAAGAGCTGTTAAATATTATTAAGTGAATCATCTCCATTAAAACTTTAG 3574
3318 tgaagatcaatccaagaagaataattatgcagctttagaaaacttlaaaatltag 3377
|||||
3575 TGAAGATATCAATTTCAACAGCAAGATTAATATGCGACTTTAGAAAACTTTAAAGTATTAAG 3634
3378 taaattagaaggaaaatlaaaggataattlaaattagaaaaagaanaatlacatctt 3437
|||||
3635 TAAATTAGAGAGAAATTAAGAGATTAATCTAAATTTAGAAAAAGAAAAATATTATCATCTT 3694
3438 atcaagtgatcatcatcttaattgtcgaatlaaagaagtaataaaaataaatta 3497
|||||
3695 ATCAAGAGGTTTACATCAATTTAATTCGTAAATTAAGATTAATTAATTAATTAATTA 3754
3498 taagaatcttcccaagtgaaaataataggaatgtaacaaatgcatagaatcttagaa 3557
|||||
3755 TACAGTTAATTTCTCCAGCGTAATATATGCGATGTTAACATGCAATTAAGATCTTACAA 3814
3558 aaattctcccaagaagacagatgttgaacagttgtaagtgaaaagtgatccgacac 3617
|||||
3815 AAAATTTCTCCAGAGAAAGAGATGTTGCAACAGTTGTAAGTGAAGAAAGTGATCCGACAC 3874
3618 attagaacaagtcacaaagaagaacagacatcaatctgttagagcagagcttcaaac 3677
|||||
3875 ATTAGACCAAAAGTCACAAACCAAGAACACATCACTGATGAGAGCAGACTCTAACAC 3934
3678 aataaacaatcaaaaatgttgaatgaatgaatgaatgaatgaatgaatgaatgaatga 3737
|||||
3935 AATTAACAAATCATCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3994
3738 tggagaatccgaagaagaatltatgatatttagaacaagtagtaacaggaagaacaaat 3797
|||||
3995 TGGAGATCCGAAGAAGATTTATGATGATTTAGACAAAGTAGTAAACGAGAACACATTAAC 4054
3798 tcccttcgtaattgataacatacttctaaatltgaatgaatgaatgaatgaatgaatga 3857
|||||
4055 TACTTCCGTAATGATTAACATCTTCTTAAATTTGAAATTTGAATTTAGATTTATATT 4114
3858 aaaaccttgaagagtggtttagaagtttaaaaaaacatctagaanaataagttatgac 3917
|||||
4115 AAAACCTTTAGAGAGTGTATTAGAGTTTAAAGAAATTAAGAAATTAAGAAATTAAGAA 4174
3918 attaatgtlaatgtttaaagatattttaaattcaacgatttaaagaagtgaaatttcaa 3977
|||||
4175 ATTTAATGTTAATGTTAAGATATTTTAAATTCACGATTTAATTAACGTAATTTTCAA 4234
3978 aaatgtttagaagatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 4037
|||||
4235 AATGTGTTAGATCATGATTTAATTCATTAAGATTAAGATTAAGATTAAGATTAAGATTT 4294
4038 caaagatccaataataatcttaataaagaagaagaagaagaagaagaagaagaagaaga 4097
|||||
4295 CAAAGATCCATTAATTAATTTCTTAATTAAGAAAAAGATTAATTTCTTAAGCAAGTTATTA 4354
4098 ttaataaagatccaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 4157
|||||
4355 TTATATTAGAGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4414
4158 taaatataatccgaanaataaatacagaatctgaatcaatcaatcaatcaatcaatca 4217
|||||
4415 TAAATATATTATCCGAAAAATTAATTAACATGATTTAGATTAATTAATTAATTAATTA 4474
4218 caaagaagtgaaaatgaagaataatcttcccttttaacaataatgaagaacctatata 4277
|||||
4475 CAAACAGAGTGAAATGAGAAATACCTTCCCTTTTAAACAATATTGAGACCTTATATTA 4534
4278 aacagcttaatgataaataatgattattgttaattcaattagaagaagaagaagttcatt 4337

|||||
Db 4535 AACAGTTAATGATTAATTAATGATTTATTGTAATTCATTAGAGCAAAAGTTCTTAATTA 4594
4338 tatataggaataatcaaacgttagaagtttaaataaagaacttaattactttaaacaat 4397
|||||
Db 4595 TACATATGAGAAATCAAAAGCTAGAAAGTTAAATTAAGAAACCTTAATTTACTTTAAACAAAT 4654
4398 tcaagaacaatltggcagatcttlaaaaaaaatacaaatcttcgttggaattgctattac 4457
4655 TCAAGCAAAATTTGGCAGATTTTAAAAAAATTAACAATTTGTTGGATTCCTATTATTC 4714
4458 aacagatlaaacacataaactaattatgacaagttccttagttagcagatagttttga 4517
4715 AACAGATTAATTAACCAATTAATTAATTTAGCAAAAGTTCTTATGACAGATGATGTTTGA 4774
4518 aaatctgtcaaacggttttacttaattacttga-tggaaactgtcgaagctatgaa 4576
|||||
Db 4775 AAATCTGCTAAATCCGTTTATCTTAATTTACTTGAATTTGGAACCTTGCAAGTATGTTAA 4834
4577 acattcaacaacacaaatggtlaaaaaaacaatg--tccacaatcttgcagatgttcaag 4634
4835 ACATTTCAACACCAATGCGTAAATAAACAATGATCCAAACAGATTTCTGATGTTTCAAG 4894
4635 acatttagatgaagaagaagaatgtlaaatgttattlaaattacaacaagaaggtgataa 4694
4895 ACATTTAGATGAAGAAGAGAAATGTAATGTTTATTAATTAATTAACAAACAGAGTATGTA 4954
4695 atgtgttgaataatccaaatctactctgttgaagaataatgtgtgagatgagatgc 4754
4955 ATGTGTTGAAATTTCAAAATCTCTCTGTAACGAAATTAATGAGTGATGATGATGATGAT 5014
4755 caaatgacgaagaagaatcagtagcaaggaagaagaataacatgtaagtactaa 4814
5015 CAATGATCCGAAGAAGATTCAGTAGCAACGGAAGAAATACATGCAATGCTCTTAA 5074
4815 acctgattcttccacacttctcagatgttattcttgcagttccctcaacttcttagaat 4874
5075 ACCTGATTTGATTCACACTTTTCGATGTAATTTCTGACAGTCTCTCAACTCTTGAAGAT 5134
4875 atcaatcttaataataatcaatgtlaaattatatacagttcaattaa 4920
5135 ATCATTTTATTAATTAATCTCATGTTAATTAATTAATTAATTAATTAATTAATTAATTA 5180
RESULT 6
PEP190 5282 bp DNA INV 21-FEB-1996
LOCUS P.falciiparum gp190 (MSAI, MSPL, PMMSA) gene for merozoite surface
DEFINITION antigens precursor.
ACCESSION X03371
VERSION X03371.1 GI:929797
KEYWORDS gp190 gene; merozoite surface antigens; MSAI gene; MSPL gene; PMMSA
SOURCE gene.
ORGANISM malaria parasite P. falciparum.
Plasmodium falciparum.
REFERENCE 1 (bases 1 to 5282) Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS Mackay,M., Goman,M., Bone,N., Hyde,J.E., Scaife,J., Certa,U.,
Stuenkelberg,H. and Bujard,H.
TITLE Polymorphism of the precursor for the major surface antigens of
Plasmodium falciparum merozoites: studies at the genetic level
JOURNAL EMBO J. 4 (13B), 3823-3829 (1985)
MEDLINE 86136024
REFERENCE 2 (bases 1 to 5282)
AUTHORS Pan,W., Tolle,R. and Bujard,H.
TITLE A direct and rapid sequencing strategy for the Plasmodium
falciparum antigen gene gp190/MSAI
JOURNAL MOL. Biochem. Parasitol. 73 (1-2), 241-244 (1995)
MEDLINE 96123395
REFERENCE 3 (bases 1 to 5282)
AUTHORS Bujard,H.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-1995) H. Bujard, Zentrum fuer Molekulare Biologie

der, Universitaet (ZMBH), Im Neuenheimer Feld 282, Heidelberg
69120, FRG
On Aug 5, 1995 this sequence version replaced gi:9920.
Location/Qualifiers
FEATURES
source

1..5282
/organism="Plasmodium falciparum"
/isolate="K1"
/db_xref="taxon:5833"
/chromosome="9"
/dev_stage="blood stage"
72..128
/gene="gp190, MSAL, MSP1, PMMSA"
/product="putative"
72..4964
/gene="gp190, MSAL, MSP1, PMMSA"
72..4964
/gene="gp190, MSAL, MSP1, PMMSA"
/codon_start=1
/product="precursor for the major merozoite surface
antigens"
/protein_id="CAA27070.1"
/db_xref="GI:929798"
/db_xref="SWISS-PROT:P04932"

gene
CDS

translacion="MKITFLCSFLFEIINTOCVTHESYQELVKLEALEDAVLGYS
LEHKEKMLNEBEIITTKGASQSGTSGSPSGSPSGSPSSRSNTLPSSNTSGA
SPPADASDASADYADLKHRYNYLTITELKYLPDLTNHMLTCDNIHGFYLD
GYEINELLYKLNFYEDLRLAKLNDYCAIDCOIPENKIRANEDLVKLVGYRKP
LNIKNQNGMEDYIKKNTKTENINELIESKKTINKKPPRANSNTPTLDDK
STYKQLEPAHNLISYLEKRIIDTLKKNIEKLELDKINKKPPRANSNTPTLDDK
NKKIEBEKEIKETIAKTIKFNIDSLDLELELYLREKNKIDISAKETEESTEPN
EYPCNVYPLSYNDINNALNELNSFGDLINPDYKPEKPNITDNEKMFINEKPN
IKIEKKIESDKSYEDRSKSLNDITKEEKLINELYSKPNINIDLPNEKMGKRY
SYKVELEHNTFASYSKHNLEKLTAKLYMEDSLBNIVYEKLYKKIKISKE
NIETLEVENIKKDEOLPEKRTKDEKNDEKLEISYDQVVOYOKVILMNKIDELK
TOLLINKVELKHNHVPNSYKQENKOEPIYLYIKETDKLVPAKPVESLNEEKN
ITTECOSDSESTEGEITGOATTKPGQOAGSALBDSVOAOAODEKQAPVPAVP
EAKAOVPTPAVNNKTEVNSKLDVLEKLYEFLNLSYICHRKILVSHSTNMRILKQY
KITKEESKLSGCDPLDLFNIONNIPWYSMFDSLNSYLQKILVSHSTNMRILKQY
LKDNDRKINLLEAKKYSTVSKLSSSQPLSLPODPEVSANDTSHNLSNLS
KLEENILSGKNKINYOELIGKSSSENEFKILKSDPYNSFPNFKVSKNDINSL
NDESKRKLEEDINIKRLKLOLSPDLNYSKIKRILPEKPKKTKVCKKQKIDKTLIK
EOLSKLSLNPKHYLVONFVFNNKKAETAELENTLENTILKIKHKGVLKTYNG
ESSPLKTLSEESIQTEDNYSLENEFKVLSLEKGLKNDINLEKRLSYLSSGHLIA
ELKEVITKNNTGNSPENNNTNNALEBYKFLPEGDVATVVSSESDTLEOSQPK
KPASTHGAESNTITTSQNVDEVDVILVIFEGSEEDYDGLQGVTEAATPSVID
NLISKIENIEVLYLKLAVYRSLKOLENNWFMENVVVKOILMSREKREMFNVL
ESDLIPYKDTSSNVVDPYKFLNKKERDKLSSYNIKSDIDPIDINANVLYGK
ILSEKYSDDLISIKRYINDKQENKYLPEFLNNITLYKTVNDKIDLEVILHEAVLN
YTERKSNVEVKIKELNYLKTIDDKLADPEKNNFVGIADLSTYHNHNNILTFELSGM
VEENLAKTVLSNLDGNLQGMINIISOHCVKKOCQNSGCFPHLDREBCKLILNKO
EDDKVENNPNTCENNGCGDADACTEEDSGSNGKIKITCECTKPDSPYLPFGICSS
SNFLISFLILMLILYSFI"

BASE COUNT 2363 a 608 c 676 g 1635 t
ORIGIN

Query Match 96.6%; Score 4750.4; DB 96; Length 5282;
Best Local Similarity 98.1%; Pred. No. 0;
Matches 4827; Conservative 0; Mismatches 66; Indels 27; Gaps 1;

QY 1 atgaagaatacattcttattgltcaattcttctttattataaatacacaatgltga 60
DB 72 ATGAAGATCATATTTCTTTATGTCATTTCTTTTATTATTAATAATACACATGTGTA 131

QY 61 acacatgaagaatgltcaagaactgtcacaataactagaagctttagaagatgtagtattg 120
DB 132 ACACATGAAGAAGTATATCAAAACCTGTCAAAAGCTTTAGAGATGAGATATTG 191

QY 121 acaggtatagattattcttcaagaagaaaaaagttaataataagaagaagaatgtagaaca 180
DB 192 ACAGGTATAGATTATTCTTATTAAGGAAAAATGATCTTAATAATGAAGAGAAA----- 243

QY 181 gctgttacaactaglacacctgltcaagaaggttcaggtcgttcaggtggttcaggtgagc 240

DB 244 -----TTACTACAAAAGGTGCAAGTGCTCAAGTGCTCAAGTGCTCAAGTGCT 284

QY 241 tcaagtgctcaggtggttcaggtggttcaggtggttcaggtggttcaggtggttcaggt 300
DB 285 ACAAGTGCTCAAGTGCTCAAGTGCTCAAGTGCTCAAGTGCTCAAGTGCTCAAGTGCTCA 344

QY 301 gcttcaggtggttcaggtggttcaggtggttcaggtggttcaggtggttcaggtggttc 360
DB 345 ACTTACCTGCTCAAGTGCTCAAGTGCTCAAGTGCTCAAGTGCTCAAGTGCTCAAGTGCT 404

QY 361 gatgctaaatcctcaggtggttcaggtggttcaggtggttcaggtggttcaggtggttc 420
DB 405 GATGCTAAATCTTACGCTGATTTAAACAGAGAGTGAAGTAAATCTTACGCTGATTTAA 464

QY 421 gaactcaataatcctcaggtggttcaggtggttcaggtggttcaggtggttcaggtggttc 480
DB 465 GAACGTAAATCTTACGCTGATTTAAACAGAGAGTGAAGTAAATCTTACGCTGATTTAA 524

QY 481 attcaggttcaaatatttcaaatatttcaaatatttcaaatatttcaaatatttcaaatatt 540
DB 525 ATTCAATGCTTTCAAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 584

QY 541 ttaactttatttgaattatataagaacaaataaataatgaatgtagttagttagttagttagt 600
DB 585 TTAACCTTTATTTGATTTATTAAGAGCAAAATTAATGATGATGATGATGATGATGAT 644

QY 601 tgcataatccttcaactcaactcaactcaactcaactcaactcaactcaactcaactcaact 660
DB 645 TGCAATATCTTCAATCTTAAATTCGCAAAATGAAGTAAACCTTACTTAAATTAACCT 704

QY 661 gttctggatataagaacacatataagaacacatataagaacacatataagaacacatata 720
DB 705 GTTCTGGATATAGAAACCATATAGCAATATTAAGTAAATAGTAAATAGTAAATAGTAA 764

QY 721 tacattaaaaaaataaaaaaacatagaataataaataatgaatgtagttagttagttagt 780
DB 765 TACATTAAAAAAATTAATAAACCATAGAAATTAATTAATTAATTAATTAATTAATTA 824

QY 781 aaaaacattgataaaataagaatgtagttagttagttagttagttagttagttagttagttagt 840
DB 825 AAAACATTGATTAATAATTAATAAGTAAACATGAAGAGAGAGAGAGAGAGAGAGAGAG 884

QY 841 gccaatatgattcttctatcttcaataaacaattagaagaacacataatataagaac 900
DB 885 GCCTATATGATCTTCTATTTTCAATTAACATTAAGAACACATTAATTAATTAAGC 944

QY 901 gtttgaagaacgtatgtagcatttcaaaaaaaatgaagaacatlaaggaatattgtagt 960
DB 945 GTTTGAGAAAAACGTATGTGACACTTTAAAAAAAATGAAGAACATTAAGGAATTAAGT 1004

QY 961 aagataaataagaatlaaaaaatccccccacggccaatttcggaataatcaccaatattctc 1020
DB 1005 AAGATAATGAATTAATAATTAATAATCCCGCCCGCAATTCGGAATTAACCAATTAAT 1064

QY 1021 cttgataagaacaaaaaaatcgaagagacagaaagaagaataaagaagaatctgccaact 1080
DB 1065 CTTGATTAAGACAAAAAATTCGAGGACACGAAAAAGAAATTAATAAATAATTCGCAAA 1124

QY 1081 attaaatttaataatgtagttagttagttagttagttagttagttagttagttagttagttagt 1140
DB 1125 ATTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1184

QY 1141 gaaaaaataaataatgtagttagttagttagttagttagttagttagttagttagttagttagt 1200
DB 1185 GAAAAAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1244

QY 1201 aatgaatcacaatgtagttagttagttagttagttagttagttagttagttagttagttagt 1260
DB 1245 AATGAATATCAAAATGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1304

QY 1261 aatgaatcaatcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1320

Dh 1305 AATGAACTTAATTCCTTTGGTGATTTAAATTAATCCATTGATTATACAAAGAACCAAGT 1364
Qy 1321 aaaaacatactatcgtatgataatgaaagaaaaaaatcataatgaaatlaagaaaaaatc 1380
Dh 1365 AAAAACATATATACGTATATACAAAGAAAAAATTCATTAATGAAATTTAGGAAAAAATT 1424
Qy 1381 aaatagaaaaaaaatlaatcgtatgaaatlaaaatcttaagaaagacatctaaagtc 1440
Dh 1425 AAAATAGAAAAAATAATGAAATCTGATTAATAAATCTTACGAACACAGATCTAAGTCT 1484
Qy 1441 ttaaatgataacaaagaaatlaaaaaatlaacttaagaaatlaagatagaaatc 1500
Dh 1485 TTAATAATGATTAACAAACAAATATGAAAAATTAATGAATTTATGATACCAATTC 1544
Qy 1501 aataataatagatttaacttaacttgcgaaaaatgattggtaaaagatattcataaa 1560
Dh 1545 AATTAATTAATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1604
Qy 1561 gtctgaaacacttacacacataactcttgcacccatgaaatcttaacaataactc 1620
Dh 1605 GTTGAACAACTTACACACATTAATCTTTTGCACTCTATGAATAATCTTAACATTAATCTT 1664
Qy 1621 gaaaagttaacaaagctcttaataatataatgaaagatlaactttaaagaaatagtagt 1680
Dh 1665 GAAAGTTTAAACAAAGCTTAAATTAATGAAGATTAATCTTTAAGAAATATAGTAGTT 1724
Qy 1681 gaaaaagaaatlaaaatattataaaaaatlaaataagaaaaatgaaatlaagattgaaca 1740
Dh 1725 GAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1784
Qy 1741 ttagttgaaatataaataaagaatgaaagacgctcttgcgaaaaaaaatlaactaaagac 1800
Dh 1785 TTAGTTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAGAC 1844
Qy 1801 gaaaaataaaccgatgaaanaaatlttaagaatcctgcacatgtaaaagtaacagttcaa 1860
Dh 1845 GAAAAATTAACCGATGAAATAATTTTGAAGATCTGACATTTGTAATAAGTACAAAGTCCAA 1904
Qy 1861 aaagtcttaataagaaaaatlaagcgaattaaaaaagacccaatggttttaaaaaaat 1920
Dh 1905 AAGATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1964
Qy 1921 gtaagaatlaaaacataataatacatgctcccaatctcttaacaacaagaataaagcaaga 1980
Dh 1965 GTAGATTAATAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2024
Qy 1981 cctattatlaactatgctgctgtaaaaaaagaatctgaataatlaaagtgctcatgctaaag 2040
Dh 2025 CCTATTATTATTATTGCTGTTGAAAAAAGAAATTTGATTAATTAATTAATTAATTAATTAATTA 2084
Qy 2041 gtaagaatctatgaataatgaagaaaaaaaacataaaaaagcgaaggttaactcggaata 2100
Dh 2085 GTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2144
Qy 2101 tctgaaaccaatcaacggaaagaaataaacaaggacaagcaacttaaaaaactggaacaaca 2160
Dh 2145 TCGGAACCATCAACCGAAGAGAAATTAACAGGCAAGCAATTAACAAACCTGACAAACAA 2204
Qy 2161 gcaagatctgctcttgaagagagatctagtaagaagcaagcacaagaacaaaaaagaaga 2220
Dh 2205 GGAGGATCTGCTTTAGAAAGGAGATTCAGTAACAAAGCAACAAACAAACAAACAAACAA 2264
Qy 2221 caaccacccagtaacagtaacagtaacagtaacagtaacagtaacagtaacagtaacagta 2280
Dh 2265 CAACCAACCAAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 2324
Qy 2281 ccaagtaataataaactgtaaaatgcttccaaatlaagattatcttgaataatataatga 2340
Dh 2325 CCAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2384
Qy 2341 tttttaaatatcatatataatgtaacaaatataatgtggtttcacaccacactaagaac 2400
Dh 2385 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2444

Qy 2401 gaaaagatatlaaaacaatalaaatlaacaaggaagaaagcaaatlaagttcatgt 2460
Dh 2445 GAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2504
Qy 2461 gatlcaatgaacttattgtttaataatacaaaaataacatccgtlaaigtatctatgctt 2520
Dh 2505 GATCCATTGACTTATTGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2564
Qy 2521 gataagcttaacaaatagtttatacaaacatcttaatgaaatcttaatgaaaaagaatggtt 2580
Dh 2565 GATAGCTTAACCAATGTTTATCAACATTAATTAATTAATTAATTAATTAATTAATTAATTA 2624
Qy 2581 tgaattctataaacttlaaggaataatgaaataatlaaaatcttaaggaaggaaga 2640
Dh 2625 TGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2684
Qy 2641 aaagtatccacatctgtaaaaaactcttcaagttcatcaatgcaacatlatatcaata 2700
Dh 2685 AAGTATCCACATCTGTAAAAACTCTTTCAAGTTCAATGCAATGCAATGCAATGCAATGCAAT 2744
Qy 2701 cctcaggaataaacggaaagtaagtgcaaatgatacatcaatctcacaatcttgaat 2760
Dh 2745 CCTCAGGATTAACCCCAAGTAAGTCAAAATGATGATACATCAATTCCTCAAAATTTGAAT 2804
Qy 2761 aatagtttaaatatttgaacaacataatgaagctctgggaaaaaacaataataccaa 2820
Dh 2805 AATAGTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2864
Qy 2821 gaattlaatagtcataaaagtagtgaaaactcttaatgaaagaatctlaaaagaatgtagt 2880
Dh 2865 GAATTAATAGTCAAAAAAGTAGTAAGCTTTATGAAGAAAGATTAATTAATTAATTAATTAAT 2924
Qy 2881 acatttataatgaatccttcttaacaatcttgtaaaatctaaagctgaatgatatca 2940
Dh 2925 ACATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2984
Qy 2941 tgaatgataatcaaaaagaaagaaatlaagaaagatataatlaaaataaaaaaact 3000
Dh 2985 TTGAATGATGAATCAAAAAGAAAGAAATTAAGAAAGATTAATTAATTAATTAATTAATTAATTA 3044
Qy 3001 ttaacgtatcatcttgaattatataataataataaataaaatlaagaagatctatgtat 3060
Dh 3045 TTACAGTTTATCAATTTGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3104
Qy 3061 aaaaagaacacgctggttaataataatgcaaaatlaaaaaactctctataaaaaaga 3120
Dh 3105 AAAAAAAGAAACAGTTGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3164
Qy 3121 caattagaatcaaaatlaagaatcaatlaaaccaagaacatglatlaaaaaactttct 3180
Dh 3165 CAATTAAGATCAAAATTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3224
Qy 3181 gttctcttaacaaaaaagaagctgtaaatlaagcaaacctgtaaaacacatlagaaaac 3240
Dh 3225 GTTTCTTTTAACAAAAAAGAAAGCTGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3284
Qy 3241 acaaaaatattatgaaacataataaaggaactggttaaatatataatggttgaatcact 3300
Dh 3285 ACAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3344
Qy 3301 ccaatlaaaacttlaagtaagaatcaatcaaaagaaagaaatlaatlaagcaatlaaga 3360
Dh 3345 CCATTAATAACTTAAGTGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3404
Qy 3361 aacttlaaagtaataatagaaggaataatlaaagataatlaaattagaagaag 3420
Dh 3405 AACTTTAAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3464
Qy 3421 aaaaaatlatcatactataacagtgatatacatattatgctgtaataatlaaagaagta 3480
Dh 3465 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3524

QY 3481 ataaaaataaaatatacaggtaattctccaaagtgaataatacga tgttaacaat 3540
 |||
 Db 3525 ATAAAAATTAATTAACAGTAATTTCTCAAGTGAAAAATATACGATGTTAAACAT 3584
 QY 3541 gcatlgaatcttacaataaaattctccagaagaacagatgttgcacagttgtaagt 3600
 |||
 Db 3585 GCATTAGAACTTACAAAATTTCTCCAGGAAGAACATGTTGCAACAGTTGTAAGT 3644
 QY 3601 gaaagtgatccgacacattagacaacaaagtcaaccaagaacagatcaactcagt 3660
 |||
 Db 3645 GAAAGTGATCCACACATATTAGAACAAAGTCAACCAAGAACACGATCACTCATGTA 3704
 QY 3661 ggaagcagtgctacaacaatacaacatacacaatgtcga tga tgaag tga tga cgt 3720
 |||
 Db 3705 GGAGCAGAGCTTAACACAAATTAACACATCAACAAATGTCATGATGAGTAGAGTGA 3764
 QY 3721 atcatagtlacctatatttgaagaatccgaagaagatga tga tga attaggaagtagta 3780
 |||
 Db 3765 ATCATAGTACCTATATTTGGAGAAATCCGAAGAAATATATATATATATGACACAGTAGTA 3824
 QY 3781 acaggaagacagtaactcctccgttaattgtatacacttcttaaaattgaagaatga 3840
 |||
 Db 3825 ACAGGAGAACGACTAATCTCTCCGTAATGATACATCTTTCTTAATTAATTAATGA 3884
 QY 3841 tatgaagtttataatataaaacctttagcagtgcttatagaagtttaaaaaacaata 3900
 |||
 Db 3885 TATGAGGTTTATATTAATTAACCTTTAGCAGGTGTTATGAGGTTTAAAAAACATTA 3944
 QY 3901 gaaataaagctatgacatttaattgttaagtgaagatatttaaaattcagatlaac 3960
 |||
 Db 3945 GAAATAACCTTATGACATTTAATGTTAAAGTTAAGATTTTAAATTCACATTTAAT 4004
 QY 3961 aaacgtgaataattcacaataatgtttagaagatatttaattcacaataaagtttaaca 4020
 |||
 Db 4005 AAACGTAATTTCAAAATGTTTGAATGATTAATTAATTAATTAATTAATTAATTA 4064
 QY 4021 tcaagtaattatgttgcagaagatccatataaaattcttaataaagaataaagataaa 4080
 |||
 Db 4065 TCAAGTAATATGTGTGCAAGATCCATATAAATTTCTTAATAAGAAAAAGATTA 4124
 QY 4081 tcttaagcagttataatataatgaagatcgaatgaagatcgaatgaatgaatgaat 4140
 |||
 Db 4125 TTCTTAAGCGTATATATATATTAATTAAGATTAATTAAGATTAATTAATTAATTA 4184
 QY 4141 gatgtcttgagatataataatataatataatcgaataaataaatacgaatgaatgaat 4200
 |||
 Db 4185 GATGTCTTGATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4244
 QY 4201 aaaaaataatacaacgacaacaaagg tga aat tga aat tga aat tga aat tga aat 4260
 |||
 Db 4245 AAAAAATATATCAACGACAAACAAAGTGAAATGAGAAATACCTTCCCTTTTAAACAT 4304
 QY 4261 attgagacctataataaacaagttlaa tga aat tga aat tga aat tga aat tga aat 4320
 |||
 Db 4305 ATTGAAACCTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4364
 QY 4321 gcaaaagcttaataatataatataatataatataatataatataatataatataat 4380
 |||
 Db 4365 GCAAAAGTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4424
 QY 4381 aattacttaaaaaacaatcaagaacaattgagcagattttaaaaaaaatacaaatctcgt 4440
 |||
 Db 4425 AATTACTTTAAAAACATTCAGACAAATGCGAGATTTTAAAAAAATTAACATTTTCGTT 4484
 QY 4441 ggaattgctgattatcaacagatttaacacataaactatttga caa agt tcc tagt 4500
 |||
 Db 4485 GGAAATGCTGATTTATCAACAGATTTAATCAATTAATTAATTAATTAATTAATTAATTA 4544
 QY 4501 aaggaatgatttttaaaatctgttaaaacggtttatataatctacttga tga aac 4560
 |||
 Db 4545 ACAGGATGATGTTTAAAAATCTTGTAAACCGTTTATCTTAATTAATTAATTAATTAATTA 4604
 QY 4561 ttgcaagtagtataacattcacaacacaaatgctgataaaaaacaatgctccacaaat 4620

Db 4605 TTCAAGGATGTTAAACATTTCAACAACACCAATGCTAAAAAAACAAATGTCACAAAAT 4664
 QY 4621 tctgtagtlltcagacatttagatgaagaagaagaatgtaattgtttatataatcaaa 4680
 |||
 Db 4665 TCTGGATGTTTCAGACATTTTGAATGAAGAAGAAATGTAATGTTTAAATTTACAAA 4724
 QY 4681 caagaagtgatataatggttgaataatccaaatccactgtgaacgaataatga tgg tga 4740
 |||
 Db 4725 CAAAGAGTATATAATGCTGTGAATAATCCAAATCTTACTTGTAACGAATAATGCTGTA 4784
 QY 4741 tgtgatgcagatgcacaatgttaccgaagaagatcagtgtagcaacggaataacaca 4800
 |||
 Db 4785 TGTGATGCAAGATGCCAAATGATACGAAAGAAAGATTCAGGTGCAACGAAAGAAATCACA 4844
 QY 4801 tgtgatgtactaaacctgatttctatccacttctgagtgatttctgcagttcct 4860
 |||
 Db 4845 TGTGAATGTACTTAACCTGATTTCTTATCCACTTTGAGATGTTTCTCCAGTTCTCT 4904
 QY 4861 aactcttaagaatcatcattctataatataatcatgattatataatgaattcattaa 4920
 |||
 Db 4905 AACTTCTTAGAATATCATTTCTTAATTAATTAATTAATTAATTAATTAATTAATTA 4964
 RESULT 7
 AF062349 5271 bp DNA INV 09-FEB-2001
 LOCUS Plasmodium falciparum strain HN2 merozoite surface protein 1
 DEFINITION Precursor (msp1) gene, complete cds.
 ACCESSION AF062349
 VERSION AF062349.1 GI:3859844
 KEYWORDS
 SOURCE malaria parasite P. falciparum.
 ORGANISM Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 5271)
 JOURNAL Jiang, G., Liu, R., Daubenberger, C.A. and Pluschke, G.
 Sequence analysis of the MSP 1 gene of Plasmodium falciparum from
 Haian, China
 REFERENCE 2 (bases 1 to 5271)
 JOURNAL Chung Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung Ping Tsa Chih 17
 (5), 294-297 (1999)
 REFERENCE 3 (bases 1 to 5271)
 JOURNAL Jiang, G., Liu, R.Z., Daubenberger, C.A. and Pluschke, G.
 Direct Submission
 JOURNAL Submitted (30-APR-1998) Molecular Immunology, Swiss Tropical
 Institute, Socinstrasse 57, Basel CH-4002, Switzerland
 FEATURES
 source
 1..5271
 /organism="Plasmodium falciparum"
 /strain="HN2"
 /db_xref="taxon:5833"
 /chromosome="9"
 /country="China"
 /gene="msp1"
 /product="merozoite surface protein 1 precursor"
 /note="merozoite surface protein 1 precursor"
 /codon_start=1
 /protein_id="AAC72885.1"
 /translation="MKIIFLCSPLFIINTQCVTHSEYQELVKLEALDEDAVLGYS
 LEQEKAVLNEGTSQAVTSTPGKSGSGSVASGSGSVASVAVSGSVASVAVS
 GSGSRRTPNSDSDSAKSYADLRKRVONYLTTELKYPPELFDLTNNMLTCDN
 IHGRKYLIDGTEIHELKYLNFYIDLAKLNDCAVSYCOIPENLIRANEDLVK
 KIVEGYRPLFNKQVKNQVEMDYIKNNFTTANINELLEGSRTTIDQKNADNEGRK
 KLYQAVNYPLNKOLEAHNLISVLEKRIPTLKNENIKRLKLIKIDNKKPSPANGS
 NTPNLLDKNKKIEHEKEIKETIATIFENIDSLPTDLELEYLREKNKRVADYTPKS
 DDPKTSVGIPIVYPYNGIVYPLPLDINHSLAANDKNSYCDLMMNPDKENIKNIIT

[illegible]

BASE COUNT	2313 a	687 c	690 g	1581 t
ORIGIN				

Query Match	58.0%;	Score 2853.4;	DB 5;	Length 5271;
Best Local Similarity	74.9%;	Pred. No. 1.3e-256;		
Matches 3842;	Conservative 0;	Mismatches 1066;	Indels 219;	Gaps 13;

[illegible]

OY	661	gltctcgagatataagaaaaacccatagagaataatataaagaataagtcgtggaaaaatggaagat	720
Db	726	gtgttcggatgtatgaaaaacccattagacaatattttaaagattatgtgagaaaaattggcaaat	785
OY	721	tacattaaaaaaaataaagaacccatagaaaatataaatgaaatgaattgaagaagtaga	780
Db	766	tacattaaaaaaaattttaaaccacacctatgcgaattatataatgaaatttaattgaaagcaagttag	845
OY	781	aaacaattgataaaatlaagaatgcactaaagaagaagaaaaaaatlatacca	840
Db	846	aaaacaattgattcaaaaattagaattgcagattatgaagaagaaaaaaaattatraccaa	905
OY	841	gtcgaattgctcttctatttcaataaaacaaatagaagaagccataattatlaagc	900
Db	906	gctcatattatattctttttatttttacaataaacattcagaagagccatattttatgaagc	965
OY	901	gttttagaanaaaacglatcgacactttaaanaaaatgaaaacattgaagaattacttgat	960
Db	966	gttttagaanaaaacgtatttgacacttttttataaaataatgaaaacattgaagaattacttgat	1025
OY	961	aagatataatgaaatlaaaaatccccccacggccaattctcggaatataccaccaatcttc	1020
Db	1026	aagatataatgaaatlaaaaatccccccacggccaattctcggaatataccaccaatcttc	1085
OY	1021	cttgaataagaacaaaaaatcgaagaaacgaanaaagaataaaagaatctgccaact	1080
Db	1086	cttgattagaaacaaaaaatcgaagaaacgcagaaacgaagaaaaattaaagaaattgccaaaact	1145
OY	1081	atlaaatlaataltgatagttacttacttgatccactgaaatgaagataactttaaga	1140
Db	1146	atttaattttaaacttgattgattgattttatttaccgactccactgaatttgaataatttttaaga	1205
OY	1141	gaaaaaaataaaatatttgatataagtcgaaggttgag-----acaaagaatca	1190
Db	1206	gaaaaaaatttataaaagttgattgtracaccttaattcacaagaattctracgaatactgttcaa	1265
OY	1192	actgaacccaatgaatalccaaatgtagtacttactcctcttgcatataagataatac	1251
Db	1266	ataccaaaaagttcctttatccaaaatgagattgtttatccctttaccactcattgattttgat	1325
OY	1252	aatgcttta-----aatgaactaatctcttggatgatttaataatccattgat	1302
Db	1326	aatgtatttacctgcagaaattatgatataaaattctcattgngtgaattttatgaatccttatact	1385
OY	1303	tatacaaaagaaccaaaglaaanaacatataactgataatgaagaanaaaatlataat	1362
Db	1386	aaagaaaaaatttatgaaaanaattttatcagatattatgaagaaaaaattttcatttaatt	1445
OY	1363	gaaattaaagaaaaaatataaatagaanaaaaaaaattgaaatctgataaaaaatctaac	1422
Db	1446	aacatttataaaaaacaaattgtatttgaagaaaaaaacatttatccacacaaamacaaaaat	1505
OY	1423	gaagccagatcttaagctctttaaattgatatatacaaaagaatagaanaattacttlaagaa	1482
Db	1506	aaaaaaatttactgcgaagatttatgaaaactcacaanaaagcatttatgaagaaatttacttataaaa	1565
OY	1483	attatgtagcaaatccaataabaatalagatttaactaaattcogaaaaatlgatggct	1542
Db	1566	tttttttgaataaagaaatttttaatttaattttttgacaaaagattcgtagattttaaattttacgt	1625
OY	1543	aaaagattatcaataaagtttgaacacttaagccaccttaactcttgatccctgaagaa	1602
Db	1626	gcacaaatttacctattttaaattgttgaaaaaacaaagattatatttattttttcattcctcttaatt	1685
OY	1603	aattctaaacataactctggaanaagtttaacaaaagccttaaatatataatggaaattattct	1662
Db	1686	aattctgtatttatattgtgttcaaaaattttaaanaagccttttcatattcttgaaacattatttct	1745
OY	1663	tttaaggaatatagttagttgaaaaagaatataaatalataaaaaattlaaagcaaaata	1722
Db	1746	tttaacaaaaagcaattttctgaaaaaagatttttaattcattatttatatttgaanaactgccttc	1805

1723 gaagaatgagatggaacatgattggaataataaaagaagaagacgcttttgaa 1782
1806 GAAGCTGATATATAAATAATTAACAGAGAAATAAGAGTACTGAAACAAATCTGAGAA 1865
1783 aaaaaaattcctaagaagcgaataaaccgagatgaataaattttagaagatctcgacat 1842
1866 AAAAAA-----TTTAAAGAGCTAACACATTCAGCAATCTCTTCTTGAAGATCTGAAATTT 1922
1843 gaaagaatgacagttcaaaaagttttaaagaaagaaatgacgaataaagaagact 1902
1923 GTAATAATTACAGTACAAAAAGTTTATTTATTTAAAAAATAGAGACTTATAGAAAGATA 1962
1903 caattgattttaaagaatgtagaattaaacataataacatgltcccaattcttaaaa 1962
1983 GAATTTATTTTAAAAAATGCAACACTAAAAAGATGATTTCTTACCAAAATATTTATAAA 2042
1963 caagaataatgaagaagacacttatattatggtgtgtgtaaaaaagaatgtataaata 2022
2043 CCACAAAATTAACCCAGAACCATATATTATTTAAATGTTATTAATAAAGAGATTAATTA 2102
2023 aaagtgtcctgctgaagtagaatactatgataaagaagaataaaacata----- 2076
2103 AAGCAATTTATACCAAAAGTAAAGATGTTAAAGAAAGACAGCTGCTTTATCAAGT 2162
2077 aaaaagaagtcgaatcggaatactcgaaacacacgaagaagaataaacaagacaa 2136
2163 ATTACACACACTTATGTTGAGCAGCAACCACTGAAGATGGGGTCACTCCACACAC 2222
2137 gcaactcaaaactgtgacaaagcagagatcgtctttagaagagatcagttacaaga 2196
2223 ACATTTATCCCAATGAGAGAAACGAGATTAACGAGAAACAGAGAAAGAGAAAGAAACA 2282
2197 caagacac-----agacaaaacaagcacaacacacgaagtagacagttacca 2247
2283 GTAGGACACACACACAGGTAACATTAACATTTGCCACACACACACATGACCCACAAA 2342
2248 gaagcaaaagcacaagtcaccaacacacacgacacgaataaataaactgaaatgct 2307
2343 GAAGTAATACTGTTGAAATTTCAATGGAACATTAAGATGATGACATTCACAACTTGC 2402
2308 tccaatataatcttctgaaataataatgaaattttaaactctatatatgctcac 2367
2403 ACATAAACAGCTTTATCTTAAGAAATTAAGATGAATTTTAACTAATATATATATGTCAT 2462
2368 aaatataatttggtttcacactcaactgaacgaagaataataaacaataaataat 2427
2463 AATATATATTTAGATCAAACTGATGATGACCAAAAATTTATAGGATATATATCTT 2522
2428 acaagaaggaagaagaacaaatgaatgacatgacatgacatgacatgacatgacatgacat 2487
2523 ACTCCAGAAAGAAAAATGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2582
2488 caaataaatacactgtaagtattctgattgtagagcttaaacaatgattatcacaa 2547
2583 CAATAATACCTACTGCTATGATTTCTTATATGATGATGATGATGATGATGATGATGAT 2642
2548 ctattatggaatctatagaagaagaatggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2607
2643 CTCTTTTTCATTTATATCAAAAGGAATGATTTATTTATTTATTTATTTATTTATTTATTT 2702
2608 gacaataatcaaatctatagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2661
2703 AATCCACATCAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2762
2662 actcttcaagttcatcatgacaaatgacatgacatgacatgacatgacatgacatgacat 2717
2763 AGTCTGGAATATCAACCGTAAATTAATTTCTGCTCAATCCGCACTCAAGTATTTCCCAAAAC 2822
2718 ----- 2717
2823 CAACAATCAAAATGCATCTCTTACCATTAACCAAAATGCTGTACTGTATCATCTGCTGCT 2882
2718 -----agtaagtgaatgatacacaatcttaacaaattgaataatgatttaaaa 2772

2883 GCTGTAGTGAAGAAAGTCAATGATCCCTTAACAGTATTGCTTTATTAACGATTTGAAA 2942
2773 ttatttgaaacatattgagctcttggaagaagaagaagaagaagaagaagaagaagaaga 2829
2943 GGTATTTGTTAGTCTTAAATCTTGAATAATTAATTAATTAATTAATTAATTAATTAATTA 3002
2830 gttcaaaaagtagtgaagaactcttaagaagaagaagaagaagaagaagaagaagaagaaga 2889
3003 TGTACACACAGATGGAATAATTTATGACATATTTTAAAAAATATGATGATGATGATGAT 3062
2890 aatgaatcttcaaaatttgtaaaatctaaagctgtagatataatcaatgaaatgaaat 2949
3063 AATGATGATTCACAAACATTCGTAATAATTAATTAATTAATTAATTAATTAATTAATTAAT 3119
2950 gaatcaaaaagaagaatgagaagaagaatataatataaataaataaataaataaataaata 3009
3120 GAAACACAAAAAATGCAATTAATGATGAATTAATAAATAAATAAATAAATAAATAAATAA 3179
3010 tcatgattctataataataataataataataataataataataataataataataataata 3069
3180 TCATTTGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3239
3070 acagttgtaataataaataatgcaaatlaaaacttcaattatlaaagaagaacaatlaaa 3129
3240 GAACCTGCCAAGACAAATGCAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATA 3299
3130 tcaaatatgaaatctcaatataaaccagaagcagtattataaactcttctgttcttct 3189
3300 TCAAAATTTGAATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3359
3190 aacaaaaaaaagaagctgtaataagcagaacggaagaacatgagaagaagaagaagaaga 3249
3360 AACAATAAATAAAGCTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3419
3250 ttatgaaacatataaagaagctgttaaatatataatggtgaaatcattccatlaaa 3309
3420 TTTATGAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3479
3310 actttagaagaagaatcaatcaacagaagaagaatataagcagtttagaagaacttaaa 3369
3480 ACTTTAAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3539
3370 gtaataagtaataatgagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 3429
3540 GTATTAAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3599
3430 tcaacttaacagtgatatacatltaattgctgaaatlaaagaagaagaagaagaagaaga 3489
3600 TCTTTCTTATCAAGTGAATGATCAATTTATTAATTAATTAATTAATTAATTAATTAATTA 3659
3490 aaaaattatacaggaattatctcaagtgaaagaagaagaagaagaagaagaagaagaaga 3549
3660 AAAAATTAATCAAGTAAATTTCTCAAGTGAATAATTAATTAATTAATTAATTAATTAAT 3719
3550 tcttaaaaaaattctccagaagaagaagat----- 3582
3720 TCTTTAGAAATTTTCTCCAGAGCAAAAGTTACAAACCTGTACTCCACCTCAACCA 3779
3583 -----gtgcaacagttgtaagtgaaagtgatccgacacatgaaacaagat 3630
3780 GATGTACTTCATCTCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3839
3631 caacagaagaagaacagcatcaactctatgtagagcagagatcaacaaataaacaataca 3690
3840 GAAACACAAATTAACCACTTCAAGCTCTTTATTAATTAATTAATTAATTAATTAATTAAT 3899
3691 caaaatgctgataatgaatgagaatgagcgtlaatcatgatacttcttgagaatccga 3750
3900 CAAATTTATGACGAAGAAGATGATTTCTTATGATTTATTAATTAATTAATTAATTAATTA 3959
3751 gaagattatgagattttagaagaagtagtaacagaagaagcagtaactcttccgtaact 3810

OY	541	ttaaacttttatcttgatttataagaagcaaaattaaatgaatgtaagtgtcattgatattat	600
Db	550	TTGAATTTCTACTTCGACCTTGCTTAAGGGCCAAACCTAATAATGACGTTCGCCCAATTACAT	609
OY	601	tgtaaatlaccccttccaactcttaaatlcgcgtcaacgtgaattgaagcttaactlaaaaaact	660
Db	610	TGTCMAATTCATTCATTTTGAAGATCGAGGCCAACGAGTGTGAGCGTATTTGAAGAAGTTG	669
OY	661	gtgttcgatatagaaaaaccattagacaatatataagaataatgtagaaaatggaaagt	720
Db	670	GTCCTCGGATNTGCCAAGCCCTTCGCACMACATCMAAGCACAAATGTGGAAAGATGTGAAMAT	729
OY	721	tacatitaaaaaaaataaaaaacocataganaatatataetgaaattatitbaaagaagtga	780
Db	730	TATTTTAAAAAGAATTAGAAAGACCCTTCGAGACATTTTACGAGCTGATCGAACAATCCAAA	789
OY	781	aaacaacttgaataaaataagaaatgcgaactaaagaagaagaaaaaaaattataccaa	840
Db	790	AAGACCATATAGCAAATAATTAGAAATGCAACCAAGSAGAAAGAAAGAAAGATTTGACAG	849
OY	841	gctaataatgatctcttctatttcaalaacaaatagaagaagcacatattattaagc	900
Db	850	GCCCAGTACGACCTGTCATCTTAACCAAAACACCTTGAAAGAACCCATPACTCTATCAGC	909
OY	901	gttttagaaaaagctatctgaacctttaaaaaaalgaaaacattaaagattacttgat	960
Db	910	GTCATCGGAAGAGCCCATTAGACACCCCTCAGSAACAATGAAATTTTCAASMAATCGTCGAC	969
OY	961	aagataaatgaaattaaaaatcccccaacgcggccaattctgasaatacaccaaatcttc	1020
Db	970	AAGATTTATGAAATTAAGAAVTCCTCCGCCACCCMACCTCGGGAAACACCCCTTMACCGTG	1029
OY	1021	cttatabaagaacaaaaaatcgcgaagacaagaaaaagaataataagaatatcgcsaaact	1080
Db	1030	CTGACACAGAACCAAGAAAGATTAGAGGAGCAGCAGCAAAAGAAATCAAAAGATGCCCAAAC	1089
OY	1081	attaaattataltgatlgaatttattactgatalcoactgaaatagaactatttiaga	1140
Db	1090	ATTAGATTCACAAATAGATGATTCCTCTTTACTAGATCCCCTTGACCTGAGTACTACTTGAGA	1149
OY	1141	gaaaaaataaaaatttctgataagtgaaaggttgaaacaaagaatlaactgaacc	1200
Db	1150	GAGAGCATTAAGATTAATAGACATCTCCGCCAAAGTGTGAACAAAGGAATCAACCGCAACT	1209
OY	1201	aatgaatcccnaatgygaagttaactlactcttgtcatataaogratltaacaatlyctla	1260
Db	1210	AATGAATATCCCAATGSGTGTGACGTACCCCTGTOTTTATMAGSATATCAACACGCTCTC	1269
OY	1261	aatgaacttaattctcttgttgttlaattaaaccaattgattatatacaaaagaaccgaat	1320
Db	1270	AACGAGCTCAATACCTCTCGGTGACTTGATTAAACCCCTTGATTTATPACGAAMAACCCCTC	1329
OY	1321	aaaacatatatactgtaataagaagaaaaaaattcataatgaaattlaagaagaaaaatt	1380
Db	1330	AAGAATPTCTACACAGACATGTAGAGAGAAAGAACTTTATCAAGAAATCAAGSAGAAAGATC	1389
OY	1381	aaaatagaaaaaaaattgaactgatactgataaaaaacttaccgaagacagatctaatgct	1440
Db	1390	AAAAATTTAGAGAGAAAGAAATTTGAGAGTGACACAGAAAAGTTTACGAAGACCGCAGCAAAAGT	1449
OY	1441	ttaaagtataaacaanaagataatgtaaaaattacttaaggaaattatgatiagsgaaattc	1500
Db	1450	CTTAACCATATCTACTTAAGAGGTTTGAAAAGCTGCTTAACGCAATCTATTGATTTCCAATTC	1509
OY	1501	aataataatagattttaaactaaatttcgaaaaaatlgatlggttgaagaatatcatatalaa	1560
Db	1510	AACATATTAACTCGACCTGACCAACTTCGAGAAATATGATGGAAAAOGGATACCTTACAAA	1569
OY	1561	gttgagaacttacacacataactcttgatcctctatgaaatcttaacataactct	1620
Db	1570	GTGGAAGAACTGACACCCATTAATACCTTTGATCTCTATGAAAGAAATTCCTTAAGCATATCTTT	1629

QY	1621	gaaagatlaacaaagccttlaaatalataggaagattatctttaaagataatagta	1680
QY	1621	gaaagatlaacaaagccttlaaatalataggaagattatctttaaagataatagta	1680
Db	1630	GGAAGGCTCCAAAGCTCTTAAGATATGAGAGACTATCTCTCGGAAACTTGTGTG	1689
QY	1681	gaaaaagatlaaalatalataaaaatttaacgaacaattgaaatlgatltgaaca	1740
Db	1690	GAGAAAGAACTAAAGTATTACAGAAATCTCATTAAGTAAGTGAAGAAACAGATTCGAGAGC	1749
QY	1741	ttagttgaaatatlaaaaaagatgaagacagcttthgaaaaaaaattactaagac	1800
Db	1750	CTTGTGTGAACATTTAAGAGGATGAGAACGTTGTTTGAGAAAGATTTCAAAAGAC	1809
QY	1801	gaaaataaacccgatgaaaaaattttagaagatcttgacattgttaagtaaatlcaa	1860
Db	1810	GAAATTAACCCGATGAGAGATCTCTGGAGCTCTCCGATATTTGTTAAATCCAACTGACG	1869
QY	1861	aaagtattatagtaacaaaatlgacgaatttaaaaagacctaigtattaaaat	1920
Db	1870	AAGGTCTCTCATATACAGAGTTATGTAGTCTCAAGAAAGACTCAACTCTCTTTGAGAGAC	1929
QY	1921	gtagaattaaaacataatatatcgtlcccaattlcttaacaaacaaagaaatgaacaga	1980
Db	1930	GTGAGATTAAACATATATATACATGTCGGCATAGTTATTAACGAGAGATTAAGCAGAA	1989
QY	1981	ccttattattatgtgtgttgaaaaaagaatttgataaaattaaagtgtlcaagcttaag	2040
Db	1990	CCATCTACTCCCTATCTCTCTCAAGAAAGAGATGACAAACTGAAGGTCTTCTGCCCCAA	2049
QY	2041	gtagatcatgtatgaatgaagaaaaaaaacatlaaaaacgaagtgcaatcgagat	2100
Db	2050	GTCGAGAGCCTTATATACAGAGAGAGAGAAAGAACTTTAAACTGAAAGACAGTCAAGATAAC	2109
QY	2101	tcggaacatcaaacccgaaggagaaaaaaacaggaacgaacactaacaacactlgacaaa	2160
Db	2110	TCCGAGCCTTCACAGAGAGAGAGATTAACCCGACAGGCTACCAACCAAGCCCGACAAAG	2169
QY	2161	gcaagatctgtctttagaaggagatctcagtacaagcaagcaagaacaaacaaacaga	2220
Db	2170	GCCGGTTGAGCTCTGGAAGGCGATAGCGGTGCAAGTCAAGCAAGAGAGAGAGAGAGCA	2229
QY	2221	caacacccagtaaccagtaaccagtaaccagaagaagaagcaagaatcccaacaccacga	2280
Db	2230	CAGCCTCCAGTCCAGTGCCTTCACAGAGGTTAAAGCTCAAGTCCATACCCACACAGCT	2289
QY	2281	ccagtaaaataaanaactgaaaaatgttcccaaattagattactctgaaaaattatagaa	2340
Db	2290	CCTGTGAATTAACAGACCGAGAGATGTCAAGCAACTGGAAGTCTGAGAGGCTATATGAG	2349
QY	2341	ttttaataacttcataatatatgtlcaacaaatatattgtgttcaactcaactgaac	2400
Db	2350	TTCCGGAATTAATCTCTACATCTGCCCAAAATATATCTCTGCTCTCAAGCAACTGTGAGAC	2409
QY	2401	gaaaagattatnaaacatatnaaatttcaagaggaggaagaagaatlaagttaagt	2460
Db	2410	GAGGAAGATTCTTAACAGATTAACAGATTAACAGAGAGAGAGAGATTAATGTCTCTTGT	2469
QY	2461	gatccattagactatgtttaaataacaaaataacataccgttaagtattctatgttt	2520
Db	2470	GATCCACGTGAGACTCTGTTCATATTCAGAGAACAACTTCGCTATATGATTTCTATGTTTC	2529
QY	2521	gatagcttaaacatagttatcacactattalagsaattalagaaatlaaataatgtt	2580
Db	2530	GATTAACCTTAACAAATTTCTCTCTCAAGTGTTCATGAGAGATATATGAGAAAGAGATGTCTC	2589
QY	2581	tgtaattatataaacttaagataatgaacaaatlaaaaatttatgtggaagacgaa	2640
Db	2590	TGCAACCTGTATTAACCTAAACACAGACAGAAAGATTAAAGAACCTTCTGAGAGAACTTAAG	2649
QY	2641	aaagatcacacatcgtgtaaaaaccttcaagttcaataatgaacacatatcaattaa	2700
Db	2650	AAGGCTTCACACTCTGTTTAAACTTCTCTTTCACGCTCCATGCAACCACTGTCTCTCACA	2709
QY	2701	cctcagagtaaacccgaagtaatgtgcaaatgtatgatacatcaactlctacaatttgat	2760

Db 2710 CCTCAAGACAAAGCCGAGAGCGCTAAACGAGACACCTCTCACTCGACCAACCTTAAT 2769
QY 2761 aatagtttaaatattttgaaacatattgagcttctggaaaaacaaatatataccaa 2820
Db 2770 AACTCACTGAAGATGTTTGGAACTCTCTCTCGGCAAGATAAGAACTCTACCA 2829
QY 2821 gaattaaagtcacaaaagtagtgaacatttatagaagaattataagaagtagtgc 2880
Db 2830 GAACCTATTGAGAGAAATGTCGCGAAGCTTACGAGAGAAAGTACTGAAGACAGCGAC 2889
QY 2881 acatttataatgaatcctttatacaatttgttaaatcttaagctgataattatca 2940
Db 2890 ACATTCATTAACGAGAGCTTCACTACTGCTGAATCTTAAGCCGATATCACTCT 2949
QY 2941 ttgaatgaatgaatcaaaaaggaagaattagaagaagattataatataaaaaact 3000
Db 2950 CTTAACGATGAATCTAAACCTAAGAGCTGGAAGAGACATCAATTAAGCTGAAGACA 3009
QY 3001 ttacagttatcatttgattataataataataataataatagaagaattatlgat 3060
Db 3010 CTGCACTGAGCTTCGACCTGTAACACAGTACAAACTGAACTGGAGAGACTCTTCGAC 3069
QY 3061 aaaaagaacacagttggttaaatataaatlgaacataaaaaacttattataaagaa 3120
Db 3070 AAGAAGAGACAGTGGCAAGGTATTAAGATGACATCAAGAAAGTTGACTGCTCAAGGAG 3129
QY 3121 caattgaatcaaatattgaatcactataaaccacaaagcattatataaacttct 3180
Db 3130 CACTGTAAGACCACTCAACTACGTAACAAACCAACACACTGCTGCAAGACTCTCA 3189
QY 3181 gttcttctaacaacaaaagaagctgaaatagcagaacatgaacacacattagaac 3240
Db 3190 GTTCTTCAACAGAGAAAGAGAGAGAGATGCCGAGACAGAGAACACTCTGGAGAAC 3249
QY 3241 acaaaaattatttgaacattataaaggactgtttaaattataatggtgaatcct 3300
Db 3250 ACCAAGATTTCTTCTCAACACTCAAGAGGCTGTCAGATTAATTAAGCGAGTCTTCT 3309
QY 3301 ccaattaaaacttgaatgaagaatcaatcaacaagaagataatataatgccagttaga 3360
Db 3310 CCTCTAAGACACTCTCTCCGAGGAGACATCCAGCGAGATTAACGCGACCTCGAG 3369
QY 3361 aactttaagatttaagattgaagaagaattaaagattataattaaattagaag 3420
Db 3370 AACTTAAAGGCTCTGTCTAAGCTCGAAGGAACTGAAGACAACTGAACCTGGAGAG 3429
QY 3421 aaaaaattatcaacttaacagtgatattacatcatttaattgctgaataaagaat 3480
Db 3430 AAGAAGCTCAGCTACCTCTAGCGAGCATGACCTGATCGCGAGCTCAAGAGAGATC 3489
QY 3481 ataaaaataaaaaattatacaggaattctccaagtgaataataacggaatgttaacaat 3540
Db 3490 ATTAAAGAACAGAACTTACACCGGCAATTAACCCCAAGGAGCAATTAATACACGCTGAATAAC 3549
QY 3541 gcatgaagatttacaacaaaattctccaagaagaagacagattgtgaacagttgaagt 3600
Db 3550 GCACCTGAATCTTACAGAGATTCCTGCCGAAAGAAACAGATGTCCTGAGGTCT 3609
QY 3601 gaaagtggatccgacacattagaacaaagtcacacaaagaacacacacacacattgta 3660
Db 3610 GAATCTGGCTCGACACACTGGAGAGCTGTCACAACTTAAGAGAGCTCATCTACTATGTC 3669
QY 3661 ggaagcagatcttaacacaaatcaacaaatgltcgatgaatgaatgaatgaatgaatga 3720
Db 3670 GGAGCGGATCCAAATACCAATTAACCAATCTCAAGAACTGACAGATGAGGTGATGACGTC 3729
QY 3721 atcatagtcctatatcttggagaatccgaagaagattatgaatgaatgaatgaatgaatga 3780
Db 3730 ATCATGTGCTTAATCTTCCGCGAGAGCGAGAGAGACTTACGATGATCCTCGCCACGGGTGTC 3789
QY 3781 aagaagaagcagtaactccttcgtaattgaatgaatgaatgaatgaatgaatgaatga 3840

Db 3790 ACCGTGAGCGTGTCACTCCTCTGATTAATGATTAACATTTCTGCCAAAATCGAAGACAA 3849
QY 3841 tatgaagtttattattataaaccctttagcagtggtttatagaagtttaaaaaacata 3900
Db 3850 TACGAAGTGTCTATCTGAACCTCTGGCAGGCGGTATAGGCTCTCAAGAAACAGCTG 3909
QY 3901 gaaataacgctttagcatttaattgtaattgtaagaatgaatgaatgaatgaatgaatga 3960
Db 3910 GAAATTAACGTGTGACCTTCAATGTCMAAGGACATTTTGAAAGACCGCTTTAAT 3969
QY 3961 aaacgtgaacatttcaaaaatgltttagaatcagatttaattccatataagaatttaaca 4020
Db 3970 AAGAGAAATTTTCAAGAACGCTTGGAAGAGGAGCTTGATGCTTAATTAAGACCTGAC 4029
QY 4021 tcaagtaattatglttcaaaagatccatataaatttcttataaagaagaagaatga 4080
Db 4030 TCCCTTAACACTGTTGTCAGAGCCCATACAGTCTCAATTAAGAGAGAGATTA 4089
QY 4081 ttcttaagcagttatattatattagaatgaatgaatgaatgaatgaatgaatgaatga 4140
Db 4090 TTTCTGTATGTTACAACTTATCAAGAGACTCCATGACACGATTAATTTGCTTAAT 4149
QY 4141 gatgtcttggatattataaaaatataccgaaaaatataaatcagatttagatcaatt 4200
Db 4150 GATGTGCTGGGTATTAACAAGATCTTGAGGAAATTAACAAGCTGACCTGACCTTATT 4209
QY 4201 aaaaaatatacgaacaaagaagtgaaatgaataaccttcccttlltaaacat 4260
Db 4210 AAAAGTATATCAACATTAAGAGAGCGAGAAATTAATCTGCTTCCGTAATAC 4269
QY 4261 atggagaccttatataaaacagttatgaataatgattatttlttgaattaga 4320
Db 4270 ATGGAACCTGTTACAAAGACAGTGAACGACAAATGACCTTCTGTAATTAACCTGGAG 4329
QY 4321 gcaaaagttcttaattatacatatagaagaatcaaacgtlaagttlaaaaataaagaact 4380
Db 4330 GCCAAGTCTCTCAACTATTAAGAGAGAGCAATGTGAAGTTAAATCAAGAGAGCTG 4389
QY 4381 aattacttaaaaacattcaagaacaaattggcagattttaaanaaaatacaattcglt 4440
Db 4390 AACTTACCTAAAACAAATCAAGAGAGAGCTGCGAGATTTCAAGAAATTAATTAATTCGTC 4449
QY 4441 ggaatgctgatttatacagaatgaatgaatgaatgaatgaatgaatgaatgaatga 4500
Db 4450 GGAATGCAAGCTGTCTACCGATTAATTAACCAACAAATCTCTGCAACAAATTCGTC 4509
QY 4501 acagatgatttcttgaacaaacttgcataaacggttataatcaattacttgaatgaac 4560
Db 4510 ACTGGATGCTGTGGAAGAAACCTGCGCAAAACAGTCTGAGCAATGCTCCAGAGGCAAC 4569
QY 4561 ttgcaagttatgtlaaacatttcaacaacaaatgctlaaaaaaacaatgctcacaacat 4620
Db 4570 CTGCAAGGATGCTGTAACATCTCCAGCACCAATGCTGAAGAAACAGTCCCGCCAGAT 4629
QY 4621 tctgagtttgaagacatttagatgaagaagaatgtaaatglttataattatacaaa 4680
Db 4630 AGCGCTGTTACAGCATGAGCGAGCGGAAAGATGAAGTGTCTCTGAACTACAA 4689
QY 4681 caagaagtgatataatglttgaanaatcccaatccacttltgaacgaanaatagtgta 4740
Db 4690 CAAGAAAGAGATTAAGTGGTGGAAGAACCAACCTTACTGCAATTAATAACATATGGCGG 4749
QY 4741 tgtgatgcaagatccaatgtaacgaagaagaatccaagtaacgaagaagaagaatcaca 4800
Db 4750 TGTGAGCCGATGCTTAAGAACCGAGAGAGAGAGAGGCTTAACGGAAGAAATTAACA 4809
QY 4801 tgtgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 4860
Db 4810 TGGAGTGTACTAAGCCGACTCTATCTACCTGAGAGGAGATTTTGTGCTCAGACTT 4869
QY 4861 aacttcttgaagatatcatcttataatgaatgaatgaatgaatgaatgaatgaatgaatga 4920
Db 4870 AATTTCTGGGATCTCTCTGCTGATCTCTATGCTATCTGACAGCTTCACTTA 4929

RESULT 9
LOCUS CVEJ31294 4940 bp DNA SYN 24-FEB-1999
DEFINITION Cloning vector pBSK+-MSP-1s/FCB.
ACCESSION AJ131294
VERSION AJ131294.1 GI:4455908
KEYWORDS Cloning vector; major surface protein-1; msp-1 gene.
SOURCE Cloning vector pBSK+-MSP-1s/FCB.
ORGANISM Artificial sequence; vectors.
REFERENCE 1 (bases 1 to 4940)
AUTHORS Bujard, H.
JOURNAL Direct Submission
TITLE Submitted (14-DEC-1998) Bujard H., ZMBH - Zentrum fuer Molekulare
Biologie, Universitaet Heidelberg, Im Neuenheimer Feld 282, 69120
Heidelberg, BW, GERMANY
2 (bases 1 to 4940)
Pan, W., Rayot, E., Toille, R., Frank, R., Mosbach, R., Turbachova, I. and
Bujard, H.
REFERENCE
TITLE Vaccine candidate MSP-1 from Plasmodium falciparum: a redesigned
4917 bp polynucleotide enables synthesis and isolation of
full-length protein from Escherichia coli and mammalian cells
Nucleic Acids Res. 27 (4), 1094-1103 (1999)
99128299
FEATURES
source
1..4940
/organism="Cloning vector pBSK+-MSP-1s/FCB"
/db_xref="taxon:85834"
/locus
1..4940
/organism="Plasmodium falciparum"
/strain="FCB-1"
/db_xref="taxon:5833"
10..4929
/gene="msp-1"
10..69
/gene="msp-1"
10..4929
/gene="msp-1"
/function="surface protein"
/codon_start=1
/evidence=experimental
/product="major surface protein-1"
/protein_id="CAB36903.1"
/db_xref="GI:4455909"
translation="MKIIFLCSFLFIITQCIVHESIYQELVKRIEALDAVLGYS
LFQKEMVLNEGISTAVTTPGSKGVSAGSGSVASGSGSVASGSGSVAS
GGSGNRRTNPDSNDSADKSYADLKHVRNLLTKELKPOLDFLNMHLLDLC
HGFKLIDGEEINELLYKLNFPDLRAKLANLVGANCYCOIPMLKIRANLVDLK
KLVFGYRKPLDNKQNVGMEYIKKKNTKTINILIEESKRTIDKNNATKREPK
KLYQAYDLSYKQLEBAHNLISYEKRTIDLKKNENIKELDKINEKNPPRANG
NTPNLLDNKKIEHEKEIKETAKTINKIDSLFDPLELYLEKKNKIDISAKV
EKTESEPNYEGVTPLYNDINNALNELNSFGDLPDLYTEKPNKNIYDNRK
KINIEKEIKIEKKIESDKSYEDRSRLNDITKEYKLEINTEYKSFNNIDLTN
PEMMGRKSYKVEKLTHTNTFASYSNHNLEKALKAYMEDYSLRNIVKELKY
YKNIJSKINELIETLVENIKDEEOLFEKKITKDKMPDEKILEVDIVKVOYKYL
MKNIDELKKTQILKVELKHNHVNSTRKQEKOPYLILYKKEIDKLKPMKVE
SLINEKKNIKTEGSDNSEPTGEITQATTPQOAGSALGSGVQAQOEOQA
OPPVAPVEAAQVTPPAVNNKTEVSKLDYLEKLEFLNTSYICHKYILFVSH
NMELIKQYKIRKEESKLSCLDPLDLNINONIPVMSPELSLNSLSQLEMEYE
KEWCVLYKLKNDKIKNLLEAKKVSSTVKTLSGMSQPLSTTPDKREVASNDTS
HSTNLSKLFENLISLGNKNYIYQELGOKSSSEMYEKILKDSPTFINESTNVK
SKADINLSLDESKRRKLEIDNKLKTLQSLDLYNKYKLERLFDPKKVGYKYM
QIKKTLLEQLESKLSLNNPRAVHLYONSVFNNKKEEIAETEMTLILKHY
KGLVYVNGESSLPKLTSEESLQTEEDNVASLENFVLSKLEGLKNDLNEKKSYL
SGLIHLIAELKEVINKNYTGSPENNTDVNNALESYKFLPECTDVAATVASEGS
DTLEFOSOPKPPASTHYGASNTITTSQNDDEVDVITVPIGESESDNADLGOVYK
EAVTPSYDINIISKITENEYVLYKLAVYSLKQLENNVMTFVNNKDLINSGFN
KRENNVNLSDLIPTKDLTSSNVVVDKFLNKKRKPFLUSVYIKDSIDTDINF
ADVGLGYKILSEYKSDSDSIKKYINDQGEKYLPLNINETLYKTVNKKIDLEF
IHLKAVLVNYTEKSNVEVKIKELNLYKTIODKLADFKKNNNVGCIADLSTDYNNHL

BASE COUNT 1669 a 1156 c 1060 g 1055 t
ORIGIN
Query Match 57.2%; Score 2812.8; DB 56; Length 4940;
Best local similarity 73.2%; Pred. No. 7.5e-253;
Matches 3603; Conservative 0; Mismatches 1317; Indels 0; Gaps 0;
LTKPLSTGVNFEINLAKTVLSNLLDGNLOGMLNISOHQVKKKQCPONSGCFRHLDEREE
CKCLNLYKQEGDKCVENPPTCNENNGGCDADACKEEDSGSNGKKTICECKPDSYP
LFDGIECSSNFIPIFLILMLIYSFI"
QY 1 atgaagatcatcttccttattgaatcttccttatttattataataacaaatgta 60
DB 10 ATGAAATCATTTTCTTCCCTCTCATTTCTGTATTATCATCAATCACTAGTCG 69
QY 61 accatgaagaagtataaagaactgtcaaaaaactgaagcttaagaatgtaagatg 120
DB 70 ACCCAGGAATCCTATAGGAGCTGTTAAGAACTGGAAGCTTGAAGATGCCGTCT 129
QY 121 acaggtatagttattctcaaaagaaatgtatataatgaagaacaaatgtaaca 180
DB 130 ACCGGATACAGCTGTCCGGAAGAGAAATGTGCTGATGAATAGAGCAATGCGACG 189
QY 181 gctgttaacactagtaacactgtgtcaaaaggttcaggtcaggtggttcaggtgc 240
DB 190 GCCGTTACAAACACACACACCGCGGCTTAAAGGCTGTGCTGAGCGGCTCGG 249
QY 241 tcaagtgctcaggtggtcaggtgttcaggtggtcaggtggtcaggtggtcaggt 300
DB 250 TCTGTGGCTCTTGCGGGTTCGTCGCTCGGCGAGCGTGGCATCAGTGGCTGAGT 309
QY 301 gcttcaggtgctcaggtgaatcaagaagctacaaactcctcaagaataatcaagtca 360
DB 310 GCAGAGGCGCGGTTCGCGGAACACTGCAAGAACCAATTCATGCAACTGTACGATTC 369
QY 361 gatgtcaaatctcaggtgtgttaaaacacagagtcgaataacttgttaactcaaa 420
DB 370 GACGCCAAGTCTTACCGCCGACCTCAAGCACCGAGTGAGAAACTATCTCCATCA 429
QY 421 gaactcaaatatcctcaactctgttaactcaatcaatgttaacttgtgtgtaat 480
DB 430 GACCTGAAGTACCCACAGTTGTTCGACTCAGTAATCATATGCTGACACTGTGATAC 489
QY 481 atccatggttcaaatatcttaactgtatgtatgaagaagaatlaatgaatatataaa 540
DB 490 ATTCAATGGCTCAATATATGATGATGAGGTAGCAAGATGCAATGACCTGTCAAG 549
QY 541 ttaacttattctgatttatttaagaagcaaaatlaaagtatgtatgtcattat 600
DB 550 TTGAATTTCTACTTCGACTTGTGAAGGCCCAACTGATGACCTTGCCTCAATGACT 609
QY 601 tctcaaatccttcaatcttaaatctgcgaagaatgaatgaagacttaaaact 660
DB 610 TGTCAATTCATTCATTTGAAGATCAGACCAACGAGTTGACGATTTGAAGAAGTTG 669
QY 661 gttctgatatagaaaacccatlagacaatatlaaagataatgtagaaaaatgagaat 720
DB 670 GTCTTGGAATGCAAGCCTCTGCACACATCAAGACATATGGGAAGAATGGAAGAT 729
QY 721 tacattaaaaaaataaaaaacacacacacacacacacacacacacacacacacac 780
DB 730 TATATTAAAAAAGTAAGAAACCATTCGAAACATTAAACGAGCTGATCGAACAATCAA 789
QY 781 aaaaacattgataaaaataagaatgaacacacacacacacacacacacacacacac 840
DB 790 AAGACATAGACAAAAAATTAAGATGCAACCAAGAGAGAAAAAAGAAAGATGTGTAC 849
QY 841 gttcaatatatcttctatttcaataaacaattagaagaagcacataattataaacg 900
DB 850 GCCCAGTACGACTGTTCATCTTAACAACACACTTGAAAGAACCCATTAACCTCATCAG 909
QY 901 gtttgaaaaaagtaattgacactttaaataaaataagaagaacttaagtaattctgt 960

Dh 910 GTACTGAGAACGCGATGACACCCCTCAAGAAATGAAATATATCAAGAACTGCTGCAC 969
Qy 961 aagataaataaataaataatccccccgcgaatcttcggaaatacaccaaatctctc 1020
Dh 970 AAGATTAAATGAATTAAGATCCTCCGCAAGCCAACTCTGGGAACCCCTTAACAGCGTG 1029
Qy 1021 ctgtgaataaacaanaaaatcgagagacagaaagaataaanaaatttgcgaact 1080
Dh 1030 CTGGACAACACAGAGATAGAGAGCAGCAAGAAAGATCAAGAGATGCCCAAAACC 1089
Qy 1081 attaaatttaataatgataatgatttacttgatccacttgaaatgaatacttaataa 1140
Dh 1090 ATTAACTTCAACATAGATTCTCTTACTGATCCCTTGAGCTGAGTACTTCTTGAGA 1149
Qy 1141 gaaaaaataaataatgataaagtcgaaggttgaaacaaaggaatcactgaaccc 1200
Dh 1150 GAGAAATAATTAATTAACATCTCCGCCAAAGTGAGCAAAAGAAATCAACCGAACCT 1209
Qy 1201 aatgaataccaatgaggtacttaactcttgatataaagataatgaatgcttta 1260
Dh 1210 AATGAATATCCCAATGCTGTGACGTACCTCTGTTATACATATCAACAAAGCTCTTC 1269
Qy 1261 aatgaacttaactcttggtgataatgaatcacttgatatacaaaagaacgaat 1320
Dh 1270 AACGACCTCAATAGTTCGGTGACTTGATTAAACCCCTTGATTAAGAAAGAACCCCTG 1329
Qy 1321 aaaaacataatacgtatgaagaanaaaatcactaaatgaagaatgaagaanaa 1380
Dh 1330 AAGAAATATCTACACAGACATGAGAAAGAAAGTTATCAACGAATCAAGAGAAAGATC 1389
Qy 1381 aaaaatgaanaaaaaaaatgaatcgtataaanaatctgaagaacgaatcactgaact 1440
Dh 1390 AAAATTGAAAGAGAGAAATTTGAGACTGACAGAAAGTTACGAAGACCCGACCAAAAGT 1449
Qy 1441 ttaaatgataaacaagaataatgaanaaaatcacttaatgaanaattatgataagcaatc 1500
Dh 1450 CTAAAGCATATCATTAAGATGAAGAAAGCTGAAGAGATGATGATTCCTCAATTC 1509
Qy 1501 aataaataatgatttaacttaacttcgaaanaaaatgaggtgaagaataatcactataa 1560
Dh 1510 AACAAATTAATGACCTGACCACTTGAGAAATGATGGAAAGGTAACCTTTCAAA 1569
Qy 1561 gttggaanaacttacacacataactcttgcaactcactgaanaactcactaaactc 1620
Dh 1570 GTGGAGAAATCTGACACACATATACCTTTGCATCTATGAGAAATCTTAAGCAATCTT 1629
Qy 1621 gaaaagttaacaagaagctcttaataatgagaagatcactcttaagaataatgagtt 1680
Dh 1630 GAGAAAGCTCACCAAGCTTAAAGTATGAGAGCATATCTCTGCGAAACATTTGTTG 1689
Qy 1681 gaaaaagaattaaataatataataaataatgaagaanaaaggaatgagatgaaca 1740
Dh 1690 GAGAAAGAACTAAAGTATTACAGAAATCTATTAAGATCGAAACGAGATCGAGACG 1749
Qy 1741 ttagttgaaatataataaagaatgagaagacgtctcttgaaanaaaataactgaaga 1800
Dh 1750 CTGTTTGAACATTTAAGAGATGAAGACACTTTTGAAGAGAAAGATTAACAAAGAC 1809
Qy 1801 gaaaataaacaacgaatgaanaaaatlttagaagatcgtacatgtaagaatgacaagtcca 1860
Dh 1810 GAAATTAACCGAGATGAGAAATCTCTGAGGTCCTCGATATTGTAAAGTCGAAGTGACG 1869
Qy 1861 aagaatttatgaacaaaatgacgaatgaanaaaagactcaatgtatttaanaaat 1920
Dh 1870 AAGGTCTCTCTATGACACAGATGTGATGAATCAAGAAAGACTCAATCTTGAAGAAC 1929
Qy 1921 gtagaattaaaacataataacatgtctcccaatcttacaacaaagaanaaagaagaa 1980
Dh 1930 GTGGAGTTAAACATTAATATACATGTGCCGAATAGTTTAAGCAGAGATTAAGCAGAA 1989
Qy 1981 cctattatcattatgtgttgaaaaaagaatgataaattaaagtgctcatgctaaag 2040
Dh 1990 CATATACCTACTGCTACTCAAGAAAGATGAGCAAACTGAAGTGTTCATGCCCAA 2049

Qy 2041 gtagaattcatgtaataatgaagaanaaaaaacataaanaacgaagaagtgcaatcgaataat 2100
Dh 2050 GTGAGAGACCTGATCAACAGAGAGAAAGAACATTTAAACTGAGAGACGTAGATTAC 2109
Qy 2101 tcggaacacatcaaccgaagaagaaataaacaagacaagaactcaaaaactctgagacaaa 2160
Dh 2110 TCCGAGCCTTCCACAGAGAGATTAACCGGACAGCTACCAACCAAGCCGGACAAACG 2169
Qy 2161 gcagagcttgctttagaagagatctcaglaacaagcaagcacaagaacaaaagaacga 2220
Dh 2170 GCCGTTTCAAGCTCCAGAGCGATAGCGTGCAAGCTCAAGACCAAGAGCAAGAGCGA 2229
Qy 2221 caaccacgaatcactacacgaatcaggaagaacgaagaacgaagtcaccacacacga 2280
Dh 2230 CAGCTTCAGTGCCAGTGCCCGTTCCAGAGGCTAAAGCTCAAGTGCCTACCAACCAAGCT 2289
Qy 2281 ccagtaataataaactgaanaatgcttccaaatagattatcttgaaanaattatga 2340
Dh 2290 CCTGTGATTAACAGACCGAGATGTACGAAACTGAGACTCTTGAGAGCTATGAG 2349
Qy 2341 tttaataacttcatalatalgtaacaaatalatttgglttcacactcaactgaac 2400
Dh 2350 TTCTGAATATCATCTTACATCTGCAAAATATATCTGCTCTCAGACACTATGAAAC 2409
Qy 2401 gaaaagataataaacaataaataatcagaagaaggaagaagaacaaatlaagttcatgt 2460
Dh 2410 GAGAAATTTTAAACAGTAACAGATTAACCAAGAGAGAGAGATTAACCTTCTTGT 2469
Qy 2461 gatccatgaactcttgatttaataatacaaaataacactccttaatgtaatcactgct 2520
Dh 2470 GATCCACTGAGCTGCTGCTTCAATATCCAGAACACATTCCTTATGTTATTTGTTTC 2529
Qy 2521 gatagcttaaacaaatgattatcaacaactatltggaatttgaanaaagaatggtc 2580
Dh 2530 GATAGGCTCAACAAATCTCTCTCAACGTTTCATGAGATATATGAAAGGAGATGCTC 2589
Qy 2581 tgaatttatataaacttaagaatgataatgaanaaaatlaattatagaagagagaa 2640
Dh 2590 TGCACCTGTATTAACOTCAAGAGAACGACAAAGATTAAAGACCTTCTGGAAGAGTTAG 2649
Qy 2641 aagaatccacactgttaaaaactcttcaagttcactcaatgcaaacatlatcaataa 2700
Dh 2650 AAGTCTCCACCTCTGTAAACTCTCTTCAGGTCATGCAACACACTGCTTCACA 2709
Qy 2701 cctcagagataaacccgaagtaagtgaatgataatgataacatcactcaaaattgaa 2760
Dh 2710 CCTCAGAGACAGCCGAAAGAGAGCGCTAACGAGACACCTCTCAGCAACCTTAAT 2769
Qy 2761 aatagtttaaatatttgaaacataatgagcttggaaaaaacaanaatatataccaa 2820
Dh 2770 AACTCACTGAACCTGTTTGAACATCTGCTCTGGCAAGATTAAGAACTATGCCAA 2829
Qy 2821 gaattaatagtgtaaaaagtagtgaaactcttactgaagaagatcttaaaagtagtgat 2880
Dh 2830 GAACCTATTGAGAGAGAAATGCTCCGAGAACTTCTAGAGAGAGTACTGGAAGACAGC 2889
Qy 2881 acattataatgaatcctttacaacttttgtaaaatctaaagctgaatgataltaactca 2940
Dh 2890 ACATTTATTAACGAGAGCTTCACTACTTCTGTAATCTTAAGCCGATGATCAACTCT 2949
Qy 2941 ttagaagatgaatcaaaagaagaagaattagaagaagatataatataaanaaaact 3000
Dh 2950 CTTTAACGATTAATCTTAACCTTAAGAGCTGGAAGAGAGACATTAAGCTGAAGAGACA 3009
Qy 3001 ttaagattcatctgattatataataataataataatgaagaagatcttgat 3060
Dh 3010 CTGCAACTGAGCTTGACCTGTTACAAAGTAACTGAACCTGGAAGAGACTCTTGAC 3069
Qy 3061 aaaaagaanaacgcttggttaataataatgcaaatcaaaaactcctlatataaaga 3120
Dh 3070 AAGAGAGAGAGCTCGCAAGTATTAAGATGAGATCAAGAAAGTTGACTCTGCTCAAGAG 3129

OY	3121	caatltagatcaaaatlgaaatgaatcaatlaaaccgaacgacgtgttcaaaactttct	3189
Db	3130	cagcttgaaagcaactcaactcagaaatccgaacagctactggaaactttcca	3189
OY	3181	gtttctttaaacaataaaagaagcgtgaatltgcgaacgtgaataacacatlaagaac	3240
Db	3190	gtgttcttcacacagaaagaaaggaagccgaatgcgcgacacagaaagacacttggagac	3249
OY	3241	acaaaataltattgaaacatlabaagaacttgtlaaaltatataatlygtgaatcatc	3300
Db	3250	accaagattcttctcaaacactcaaaagcccttgcctaaattatataatggcagcttct	3309
OY	3301	ccattaaaacttgaatgtgaagatcaatltcaacagaagaataatltatgtccagttlaa	3360
Db	3310	ctctggaagactctctccggagagacatccaaacggaaagataactgaaccgcctcgag	3369
OY	3361	aacttlaaagatctaaatgaatlagagaagaaatlaaagataatttaatttagaag	3420
Db	3370	aacttcaagctcgtgttaagcttgaaggacaccttaaggacacctgaacctggagag	3429
OY	3421	aaaaaatlcatcatlcatcaaglygaatcatlcaatttaatlycgtgaattlaaagaagla	3480
Db	3430	aagaagctcacactcctctagggagactcatcacctgatgcccgaagctcaagaagtc	3489
OY	3481	ataaaaaataaaatltatacagtaattltccgaagtgaataataacagatgtataacat	3540
Db	3490	atttaagaacaaagactcacccggacatagcccaagcgaataatatacagactgtatac	3549
OY	3541	gcattgaacttccaaaataatltccccaagaagaagatgttgcgaacaglytgaat	3600
Db	3550	gcactgcaattttacagaagttccttccgaaaggaacagatgtgccactgtgtgtct	3609
OY	3601	gaaglygaatccgaacatlatagaacaaglylcaaccaagaacacagacatcaatlyla	3660
Db	3610	gaattctggctccgacacactgagagacagctgcacaaacttaagaagctatctcatgtc	3669
OY	3661	ggagcagagcttcaacaataacaacatcaacaataatgtcatgtgaagtagaagta	3720
Db	3670	ggaccgcgactcaattacaaatttaccacatctcagacagctgacagatgagctgatacctc	3729
OY	3721	atcagtagaccttaatttggagaatccgaagaagatltatgaatgtttagaagaatla	3780
Db	3730	atcatgttgcttattcttcggcggagagcgaaggaagacttaccgatgacctgcgcagctgtct	3789
OY	3781	acaggagaagcagtaactcccttcgtaattgtatacatlcttctlaaattgaaatgaa	3840
Db	3790	accggtgagcgtgctcactctctccggtatgtatcaattctccaaatgatgaagacaa	3849
OY	3841	tatgaagtttataatttlaaaaccttaagcaggtgtltatagaagttlaaaaaacaalta	3900
Db	3850	tacaaagtgctctatgtgaaacacttcggcagagcgtctatagctctctcaagaaacgtc	3909
OY	3901	gaataagctltagaacatttaatgttlaatgttlaagatatttlaaattcaacgtlaat	3960
Db	3910	gagaaattaacctgatgaccttcaattgtcaacgtgaagaaatttcttaaacagccgtttat	3969
OY	3961	aaacgtgaanaatttcaaaaaatgttltagaatcaagatlttaattccataaagaattlaa	4020
Db	3970	aagagagaaattttcaagaaacgcttggagagagacttgatttccctataaagacctgac	4029
OY	4021	tcaagtaataatgttgcataagatccatataaatttcttaataaagaanaagatataa	4080
Db	4030	ttcttcttactacacgttgtcagaagaccatatacagattcctcaataaagaacagggattaa	4089
OY	4081	ttcttcaagcagttataatlatatlaaggttcaatagatlaaggtatataattttgcaat	4140
Db	4090	ttttctgtctgtttacacttatatacaaggaactccatgacaccgatatcaattttgcctaat	4149
OY	4141	gatgttcttgatataataaaatcatlaccgaanaataataacagatltgatltcaat	4200
Db	4150	gatgtgctgggggtattatcaagatccctgacgaaaaatatacaattctgacctgtactat	4209
OY	4201	aaaaacataatcaagcaacaagaaglygaanaatlygaanaatccctcttltlaaacat	4260

Db	4210	AAAATATATACAGCATATAGCAGGCGGAGATGAAAAATATCTGCCCTTCTGAAATAC	4269
Oy	4261	atgagacccttatataaaacagttaa tga taaattgattctatctgtaattagaa	4320
Db	4270	ATGGAACCCCTGTACAAAGACAGTGAACGACAAAAATCGACCTCTTCTGAATTCACCTGGAG	4329
Oy	4321	gcataagcttctaataatacatatgagaatcaaacgtatgaagttaataaagaactt	4380
Db	4330	GGCAAGTCTCTACACATATCTTACGGAAGAGCAAAATGTGAAGTTTAAATCAAGGAGCTG	4389
Oy	4381	aattacttaaaacaattcagaacaatttgcagattcttaaaaaaaataacaatttcgt	4440
Db	4390	AACTACTCTAAAAACATCCAGACACAGCTGGCAGATTTCAAGAAAATATACATTTTCGTC	4449
Oy	4441	ggaattgctgattatcaacagattaaacataaacttatttgaacaagcttccttagt	4500
Db	4450	GGAAATGCAGACCTGTCTACCGCATTTATACCAACAACATCTCTGACCAAGTTTGTCTCC	4509
Oy	4501	acaggtatggttttggaaaactcttcttaaacocgltttcttaattacttgatggaac	4560
Db	4510	ACTGGGATGTGTTCGAAAAACCTCGCCAAACAGTCTAGGACATTTGCTCGACGGCAAC	4569
Oy	4561	tgcgaaggtatgtttaaactttcaacaacacagatggttaaaaaaacatgctccacaatt	4620
Db	4570	CTGCAGGCGATGCTGAACATCTCCACGCCACCAATGGCTACAGAACAGTGGCCCCAGATT	4629
Oy	4621	tctbgaagtcttaagaactttagatgaagaagagaatgttanaatgttataataataa	4680
Db	4630	ACGGCGTGTTCAGGCATCTGGACGAGCGCAAGAGTGCAGTCTCTCGAACTACAA	4689
Oy	4681	caagaagtgatcaaatggttctgaanaatccaactcctacttctgaacgaanaataatgttga	4740
Db	4690	CAGAGGAGGATTAAGTCTGGGAAGAACCCAAACCCCTTACCTGGAAATGAAACATATGGCGGG	4749
Oy	4741	tgtgatgcagatgaccaatgtaaccgaagaagattcaggtgacaaacggaagaataatcaca	4800
Db	4750	TGTGAGCCGATGTCTAAATGCAACCGAGAGAGACAGCGGCTCTAACGGAAGAAATACACA	4809
Oy	4801	tgtgaatgactaaacctgattcttatccactttcgaatgtatattcttgcagttcctct	4860
Db	4810	TTCGAGTGTACTTAAGCCGACATCTCTATCCACTTCGACCTTCGAGGGATTTTTCGCCACACTCT	4869
Oy	4861	aactcttagagatatcatatttatataaacacatgtaataataataacggttcaataa	4920
Db	4870	AAATTCTCGGCACTCTTCTCTGCTGATCTCATCTGATCTCTGTACAGCTTCATCTTA	4929

RESULT 10
AF218248 LOCUS
DEFINITION AF218248 5085 bp DNA INV 24-JAN-2000
Plasmodium falciparum major merozoite surface antigen (gp195) gene.
complete cds.
AF218248
ACCESSION AF218248.1 GI:6740099
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE

malaria parasite *P. falciparum*.
Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 5085)
Shan,Z.X., Yu,X.B., Li,X.R., Ma,C.L. and Fang,J.M.
Molecular cloning and sequence analysis of major merozoite surface
antigen(gp195)gene of *Plasmodium falciparum* isolate FCC1/BN
2 (bases 1 to 5085)
Shan,Z.X.
Direct Submission
Submitted (21-DEC-1999) Department of Parasitology, Sun Yat-sen
University of Medical Sciences, 74 Zhongshan Road II, Guangzhou,
Guangdong 510089, P.R.China
Location/Qualifiers
1. 5085
/organism="Plasmodium falciparum"

/isolate="FCCL/HN"
/db_xref="taxon:5833"
/country="China"
1..5085
/gene="gp195"
1..5085
/gene="gp195"
/codon_start=1
/product="major membrane surface antigen"
/protein_id="AA27526.1"
/db_xref="GI:6740100"

/translation="MKILIFLCSPLFIINQCVTHESYOELVKKLEADLVLTGYS
LEOKMYNIEGTSCTAVTSTSGSGSVASVAGSGSVASGSGNS
RTNPSDSSDAKSVADLKHVONNYITKEIKLPPLDINMLTCONIHFGY
LIDGEEINELLYKINFYDLIRAKINACANSYQIIPNLRINELDYKRIYFGY
KPLDNIKDONGKEMDYIKKNTTIANINELLESKRTIDONKADNEGKKLYQAO
YNLFYKQLOEAHNLISLEKRIIDTLKKNEMIKLBDIDIKTKDAKLTGSPNP
LPEKKEVEGHEEKIEIKIANTIKENIDSLFDPELEYLEKKNKYDVTSPDPT
KSVQIPKVPNGIYVPLPDIHNSLAADNDKNSYGLMDNDTEKINEKIITDNKE
KRIFINNIKQIIDEKNIHNTKEQNKILDEYKSKDYELLKFEYEMKRNMPDK
DVNDKIFSAKRTYVWEKKNKSSNSNYINOKLKAISYLEYSLRKISKEDF
NHYYTLKGLADIKLLEIKSEKLEKFKGLTHSANSLESDIVKIQOVKL
LIKLEIDRLKIELFNKQLODSIHVPNIYKPNKPEYVLIIVLEKVDKLEFIPKY
KDMLEKQAVLSIIOPLVAASETDEGSHSTLVSQGEETEETVGHATTYITL
PPOSPPEKVVVENSIEHKSNDNSQALITVYKLIKDEFLTGYICHKYLIVSNS
MDOKLEYNLTPREKNELSGDPLDIFNIONNTIPAMYSLYDSNNLOHIFPELYO
KEMITYLKLKEENIKKLEOKOITGTSSTSSGNTVNTAOSTHSNSQNSNA
SSTNONGVAVSSGPAVVEESHDLIVLISINDLGIYSLMLGKTVVPNLTISTT
DEKFEYENILKNDYFENDIKQEVKSNKYITGLTQKNAINDEIKLKDTQLDSF
DLUNKYLIKLDRLFNKKELGODKMOIKLILKLEKLSLNSLNPNLQNFVVF
NKKKEAEIETENTLENTKILKHYKGLVYNGSSPLKTESEIOTEDYANLEK
FRVLSKIDGLNDNLHLGKKLSFLSGHLHITELKVIKKNKNTGSPSNKKNV
EALKSYENLEPAKTYTVTPPODPVPSPLSVRYSGSGSSTKEMTOIPTSSLTTEL
QQVQSOVYDEEDSVLVPFGESENDLIDQVTEBAISVTDNLDSGNEEDY
ILPLAGVYRSLKQIEKNITTENLNDNLNSLKRKRFLVDLENDLSGNEEDY
NEVIIEDESKLINSBOKNTLKSRYKIKESVENDIKFAOEGISYELAKYKDDLES
IKVYIKKEKEPSPPTPPSPAKTDEQKSKFLPFLNIETLYNNLVNKIKDOLIN
LKAKINDCNPEKDEAHVKITKLSDKAIDDKIDERNKNDPEAKILNDPRKMLG
KLSTGIQNPNTLISKILGKIPDMKINISOHQVOKVSCNSCPFLHDERECKC
LLNKQESDKVENVNPTCNENNGCCDADACCTEEDSSNGKLTCECTKPDSTPLFD
GIFCSSNFIISLILMLILYSEI"

BASE COUNT 2244 a 671 c 678 g 1492 t
ORIGIN

Query Match 56.2%; Score 2766.8; DB 5; Length 5085;
Best Local Similarity 74.2%; Pred. No. 1.4e-248;
Matches 3800; Conservative 0; Mismatches 1087; Indels 231; Gaps 15;

QY 1 atgaagatcatctctttatgctcatctctttttatataataacacaaatgta 60
DB 1 ATGAAGATCATCTCTTTTATGCTCATCTCTTTTATATAATAACACAAATGTGTA 60
QY 61 acacatgaagatcatcaagaactctgcaaaaaactagaagctttagaagatgacatg 120
DB 61 ACACATGAAGATCATCAAGACTGTGCAAAAACTAGAGCTTTAGAACATGACAGATTG 120
QY 121 acaagttatagttatcaaaaagaaaaatggtatataatgaagaagaatgtagaca 180
DB 121 ACAGGTTATAGTTATTTCAAAAAGAAAAATGATTAATAAGAACAGAACATGTGACAA 180
QY 181 gctgttacaactagacacctggttcaaaaggttcagttgctcaggtcaggttcag 240
DB 181 GCTGTTCACACTAGTACACCTGCTTCAAGTGTTCAGTTACTTCAGGTG----- 240
QY 241 tcaagttgctcaggttcaggttcaggttcaggttcaggttcaggttcaggttcag 300
DB 241 -----TTCAAGTGTGTTCAAGTGTTCAGTTGCTTCAGGTGTTCAAGTGTGTT 282
QY 301 gcttcaggttcaggttcaggttcaggttcaggttcaggttcaggttcaggttcag 360
DB 301 GCTTCAGGTGTTCAAGTGTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGTTCAG 342

QY 361 gatgctaaatcttagctgattttaaaacagagtagcgaataacttgtaactatcaaa 420
DB 361 GATGCTAAATCTTAGCTGATTTAAACATAGAGTTCAAAATTAATCTGTCATATTAA 402
QY 421 gaactcaaatatcccaactccttgatttaactcaatcatatgtaactctgtgtgata 480
DB 421 GAACCTAAATATCCGACACTCTTGATTTAAACCAATCATATGTTACTTGTGTATAT 462
QY 481 atcaatgcttcaaatatcaatgtagatgtagalgaagaatgaatgaatattataaa 540
DB 481 ATTCATGCTTTCAAAATATTTAATGATGATATGAAGAAATTAATGAATATATTA 522
QY 541 ttaactttatattgattatataagaagaataatgaatgtagatgtagatgtagat 600
DB 541 TTAACCTTTATATTATATTATTAAGAGCCAAATTAATGATGATGATGATGATGAT 582
QY 601 tgcataatcccttcaactccttcaaaatctgcaatgaaatgtagatgtagatgtagat 660
DB 601 TGTCATAATCCTTTCAATCTTTAAATTCGTCGAATGAATGATGATGATGATGAT 642
QY 661 gttctcgatataagaacccaatagaacatattagaataatgtagaagaatgtagat 720
DB 661 GTGTTGGAATATGAGAAACCATAGACAAATATTAAGATAATGTAAGAAATGAG 702
QY 721 taacttaaaaaataaaaaaacatagaagaatataatgaatgaatgtagaagaatg 780
DB 721 TACATTAATAAAATAATAAACACATAGCAAAATTAATGAATTAATTAATGAGAA 762
QY 781 aaaaacttgataaaaaataagaatgcaactagaagaagaagaagaagaagaatata 840
DB 781 AAAACAATTGATCAAAATTAAGATGAGATTAATGAAGAAAGAAAGAAATTAAT 822
QY 841 gctcaatataatccttcttcttcaataaaacatagaagaagcacataatgaatg 900
DB 841 GCTCAATATAATCTTTTATTTAATAAACAATTAACAAAGACATTAATTAATTA 882
QY 901 gttttagaagaagaatgtagacactttaaagaagaagaagaagaagaagaatgtag 960
DB 901 GTTTTGAAGAAAGATTAATGACACTTTTAAAGAAAGAAAGAAAGATTAATTA 942
QY 961 aagataatgaatataaa---atccccaacgycgaacttctggaatatacacaat 1017
DB 961 GATATGATATAAATTAATAACAGATGCCGAAACCTCAGTACGAGTAAACCAAT 1002
QY 1018 ctccctg---ataagaacaaaaaatcgaggaacacgaaagaagaagaagaatg 1074
DB 1018 CTCCCTGAGAGATTAAGAAAGAGAGTGCAGGACGACGAGAAAGAAATTAAT 1062
QY 1075 aaaaacttaaatataatgtagatgtagatgtagatgtagatgtagatgtagat 1134
DB 1075 AAAACTATTAATTTTACATTTGATGATTTATTTACTGATCCACTTGAATTA 1122
QY 1135 ttaagaagaagaagaataataatgtagatgtagatgtagatgtagatgtagat 1185
DB 1135 TTAAGAGAAAGAAATTAATAAGTGTGATGATGATGATGATGATGATGATGAT 1182
QY 1186 gaatcaactgaacccaatgaatataccaaatgtagatgtagatgtagatgtagat 1245
DB 1186 GTTCAAAATGCAAAAGCTTCTTATCCAAATGATGATGATGATGATGATGAT 1242
QY 1246 attaaatgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtagat 1296
DB 1246 ATTCAATTAATTTACTGCTGATGATGATGATGATGATGATGATGATGATGAT 1302
QY 1297 ttgattatatacaagaagaacgaatgaagaacataatgtagatgtagatgtagat 1356
DB 1303 GATFACTAAAGAAAGAAATTAATGAAGAAATTAATGATGATGATGATGATGAT 1362
QY 1357 atcaatgaatataagaagaagaatataagaagaagaagaagaagaagaatgtagat 1416
DB 1363 ATTAATAATCAATTAATAAAGAAATTAATGATGATGATGATGATGATGATGAT 1422
QY 1417 tcttaagaagaagaatgtagatgtagatgtagatgtagatgtagatgtagatgtagat 1476

```

1423 CAAATAAATAATTAATCTTGAAGATTATGAAAGTCAAAAAAGATTATGAAGATTCTT 1482
1477 aatgaattatgatagcaaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 1536
1483 GAAAAATTTTATGAAATGAAATTTTAAATTAATTTTGCACAAAGATGTCGATGATTAATA 1542
1537 atgggtataaagatataatcaataaagttgagaacttcaacacataatcttctgcatcc 1596
1543 TTCAGTGCAGATATACATATTAATGTTGAAACCAAAAAATTAATATTAATTAATTTTCATCC 1602
1597 tatgaataatctcaacataatcttgaagaattcaacaagctcttaaatatataatgaagat 1656
1603 TCATAATTAATCTGATATTAATTAATGTTCAAAAAATTAATAAAGCTCTTTCATTAATCTTGAAGAT 1662
1657 tatctttaaagagatagatagtttgaagaagaattaaatataatataaataatcaatgaagc 1716
1663 TATTTCTTTAAGAAAAGGAATTTCTGAAAAAGATTTAATCATTAATTAATCTTGAAGAACT 1722
1717 aaaaatgaagaatgagattgaaacatagtttgaataataataaagaatgaagaacagctc 1776
1723 GGCCTCGAAGCTGATATTAATAAATTTAAGCAAGAAATTAAGAGTAGTGAACAAATTT 1782
1777 ttggaataaataatctactaagaacgaataataaaccagatgaataaatttgaagatctc 1836
1783 TTTGAAAAAATAA---TTTAAAGACGCTAACACATTCAGCAAAATGCTTCTTGAAGATCT 1839
1837 gacattgtaagaatgacaagttcaaaaagttttaaagacaaatctgacgaataaataa 1896
1840 GATATTGTAATAATTCACAGTACAAAAAGTTTATTAATTAATAAATAAAGAAAGACTTTAAGA 1899
1897 aagaacaaatctgaattcaaaaaatgtlaagaataaacaataataacatagttcccaatctc 1956
1900 AAGATGAGATTTATTTTAAATAAATGCAACAACTAAAGATGATGATGATGATGATGATGAT 1959
1957 taacaacaagaataataagaagaacattatctaatcttgaatcttgaagaagaagaatctg 2016
1960 TAAATAACCAAAAAATTAACACAGAACATATTAATTAATTAATTAATAAATAAAGAAAGATGAT 2019
2017 aatttaaagaatgtctgacgttaagaatgacatctgaataaagaagaagaagaagaagaaga 2076
2020 AATTTAAAGAAATTTTATTAACCAAAAGTAAGAACATGTTTAAGAAAGAACAGCTGCTTTA 2079
2077 aaaaacagaagttcaatcagataatctggaacacatcaacgaagaagaagaataacagagaca 2136
2080 TCAGATATTACACAACTTATGTTGACGACAGGAAAC---AACTGAAGATGGGGGTCAC 2136
2137 gcaacatacaaaacctggaacacagacagatctgctttagaagagatctgaatcaagca 2196
2137 TCACACACACACATTAATCCAAATCAGAGAAACAGAAAGAAACAGAAAGAAACAGTAGACAC 2196
2197 caagcaagaagaacaaacaaacagacacacacagtagcagttaccagttacagaagaaga 2256
2197 ACAACACACGTAACATTAACATTACACACACACAAACCATCACACCAAAAAGAGTAATAA 2256
2257 gcaacaagttcccaacacacacacagacagcaatgaataaataaactgaataatgtttcaaatca 2316
2257 GTTGTGTAAGAAATTTCAATAGACATTAAGAGATGATGATGATGATGATGATGATGATGAT 2316
2317 gattatcttgaagaataatcttgaataatcttgaataatcttgaataatcttgaataatctt 2376
2317 GTTTATCTAAGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2376
2377 ttggtttcaacacacacacacacacacacacacacacacacacacacacacacacacacacac 2436
2377 TTAGTATTCAAACTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2436
2437 gaagaagaacaaatgaatctatctgataatcttgaataatcttgaataatcttgaataatctt 2496
2437 GAAAAAATGATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2496
2497 ataactgataatcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 2556
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

3559 aaattctccagaaggaacagat----- 3582
3634 AATTTCCTCCAGACGAAAGTTTACACAGTTGTACTCCACCTCAACCAATGTAAT 3693
3583 ---gttcaacagcttglaagtgaaagtgatccgacacattagaacaaagtcaccaaag 3639
3694 CCATCTCATTTATCTGTAAAGGTAAGTGTAGTTCAGATCCCAAAAGAGAAACCA 3753
3640 aaacagacatcaactcattagagagagctcaacacataacacatcaacaaatgtc 3699
3754 ATACCACTTCAGGCTCTTATTATTAACAGAAATTAACAACAGTACTACATCACAAATTA 3813
3700 gatgagaaagtagagagcattacatcagctatattttggaaatccgaagaagattat 3759
3814 GACGAGAAAGATGATTCCTTAGTGTATTAACCATTTTGGAGAAATCCGAGATTAATGAC 3873
3760 gatgatttagaagaagtaglaacagagaagacagtaactcttcglaattgatacata 3819
3874 GAATATTTAGATCAAGTACTGAGAAAGCAATA---TCTGTCACAATGATTAATATC 3930
3820 cttctcaaaattgaaatgaatagtggtttatatttaaaaccttagcaggtgttat 3879
3931 CTTCTAGGATTTGAAATGAAATGATGTATATTAATTAACCTTAGCTGAGATATAT 3990
3880 aagaagttiaaaaaaacaattagaanaatacgttatagacattiaatgttaatgttaagat 3939
3991 AGAAGCTTAAAAAACAAATTGAAAAAACATTATTAATTAATTAATTTGAACGAT 4050
3940 attttaaattccagattataaagcgtgaanaattlcaaaaatgttttagatcagattla 3999
4051 ACTTTAAATTCACGCTTTAAGAAACGAATAATTTCTTGATGTATTAAGATGATTTA 4110
4000 attcatalaagaatttaacatcaagtaattatgtgtcgaagatccataaattctt 4059
4111 ATGCATTTAAACATATATCCCAATGAAATGAAATGAAATGAAATGAAATGAAATGAA 4170
4060 aataaagaanaagagataaattcttaagcagattataaatttaagtttaagattatagat 4119
4171 AATTGCAACAAAAAACACACTTTTAAAAAGTTTAAACAAATATTAAGATGATGAGAA 4230
4120 acgagataaattttgcaaatgagtctcttgatattataaataatlatatccgaanaatat 4179
4231 AATGATATTAATTTGACAGAGAAAGTATAGTTATTAAGTAAAGGTTTACGGAATAT 4290
4180 aaatcagatttagattcaattaaatatataca----- 4214
4291 AAGGATGATTTAGAAATCAATTAATAAAGTTATCAAGAAAGAAAGAGTTCCCATCATCA 4350
4215 -----cgacaacaaggtgaanaatgagaatatc 4242
4351 CCACCAACAACCTCTCGTACCAGCAAAAAACAGCAACAAAAAGAAAGTAAGTTTC 4410
4243 ctctcccttttaacaatatltgagacctatataaacaagtiaatgaataatgttla 4302
4411 CTTCCATTTTAAACAACATTTGAGACCTTATTAACAATTAATTAATTAATTAATTAAT 4470
4303 ttgttaattcattagaagcaaaagttcttaaatatatacatatagagaatcaacgtlaga 4362
4471 TACTTAATTAACCTTAAGCAAAAGATTAAAGATGTAAATGTTGAAAAAGATGAACACAT 4530
4363 gttaaataaagaacttaattacttaaaaaacatccaagacaacaaatgtgcagattttaa 4422
4531 GTTAAATTAACCTTAAGTAAAGCAAAAGCAATTAAGATGATGATGATGATGATGATGAT 4590
4423 aaaaataacaattcgttggaattgtgatttlaacaacagattataacataacta 4482
4591 AACACATAGACTTCGACACCAATTAATAAATTTGATTAATGATATATGCAAAAAAGATATG 4550
4483 ttgaacaagttcccttagtaacaggtatgttttgaanaatcttgtaaacggtttatct 4542
4651 CTTGGCAAAATTAATTAAGTACAGGATTAG---TTCAAAATTTTCTTAATACAAATATATCA 4707

4543 aattacttgatgaaacttgcagagtgatglaaacaattcacacaacacatgcgtaaaa 4602
4708 AATTATTAATTAAGCAAAATTTCCAAAGATATGTTTAAACATTTTCAACACCAATCGTAAAA 4767
4603 aaacaatgtccacaanaattctgtgatttcaagacatttagatgaagaagaatgtaaa 4662
4768 AAACAATGTCAGAAAATTTCTGTGATGTTTCAGACATTTATGATGAAGACAGATGTAAA 4827
4663 tgtttatlaattacaacaagaagtgatgaatgtgttgaanaatccaaatcctactgt 4722
4828 TGTATTTAAATTAACAACAAGAGGTGATTAATGTTGAAAAATCCAAATTCCTACTGT 4887
4723 aacgaataaattgtgagtgatgtagcagatgccaaatgtacccaagaagatcaggttagc 4782
4888 AACGAAATTAATGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4947
4783 aacgaagaanaatacatalgtgaatgactaaacctgattcttatacactttcagatgt 4842
4948 AACGAAAGCAAAATCAACATGTAATGATTAACACGATGCTTATCATTCTGATGT 5007
4843 atttctgcagcttcccttaactctttagaataatccttataatcattcagatgaata 4902
5008 ATTTCTGAGTTCCTTAACCTTTAGGAATATCAATCTTAATTAATCAATCAATTAATTA 5067
4903 ttatacagttcatlaa 4920
5068 TTATACAGTTTCATTTTA 5085

RESULT 11
AF062348
LOCUS
DEFINITION
Plasmodium falciparum strain HNI merozoite surface protein 1
precursor (msp1) gene, complete cds.
ACCESSION
AF062348
VERSION
AF062348.1 GI:3859842
KEYWORDS
malaria parasite P. falciparum.
ORGANISM
Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE
1 (bases 1 to 5243)
Jiang,G., Liu,R., Daubenberger,C.A. and Pluschke,G.
Sequence analysis of the MSP 1 gene of Plasmodium falciparum from
Hainan, China
JOURNAL
Chung Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung Ping Tsa Chih 17
(5), 294-297 (1999)
REFERENCE
2 (bases 1 to 5243)
Jiang,G., Liu,R.Z., Daubenberger,C.A. and Pluschke,G.
Direct Submission
JOURNAL
Submitted (30-APR-1998) Molecular Immunology, Swiss Tropical
Institute, Socinstrasse 57, Basel CH-4002, Switzerland
LOCATION/Qualifiers
1. 5243
/organism="Plasmodium falciparum"
/strain="HNI"
/db_xref="taxon:5833"
/chromosome="9"
/country="China"
/gene="msp1"
/product="merozoite surface protein 1 precursor"
/gene="msp1"
/gene="msp1"
/note="binds spectrin"
/codon_start=1
/product="merozoite surface protein 1 precursor"
/protein_id="AAC72884.1"
/db_xref="GI:3859843"
/translation="MKIIFLCSPLFIINTOCVTHESYOELVKRLALEDAVLVNGYS
LROKEMVINEGSGAVTSTPGSGSVTSVSGSGSVASVSGSGSVASVSGSGS
RTNPDSNDSDSADKSYADLKIRVQNYLFTIKELYPELFDLTNHLTLCDNIHGKX

1657 taattcttaaggaatagtagtctgaagaaagaataaataatataataaataaataaagc 1716
1728 TATCTTTAAGAAAGAAAGATTTCTGAAAAAGATTTTAATCATATTATATCTTGAAACCT 1787
1717 aaaaatagaatagtgaatctgaacatagttgaaatataaataaagaagaagaacagctt 1776-
1788 GGCCCTCGAAGCTGATATAAAAAATTAACAGAAAGAAATTAACAGTAGTGAACAAATTT 1847
1777 ttggaataaaataattacttaagaagcaaaataacacagatgaagaaaaatttgaagatct 1836
1848 CTAGAAAAA---TATTAAAGACTTAACACATTCAGCAAACTCTTCTTAAGATATCT 1904
1837 gacatttgaagtagcaagttcaaaaagtttattcaatgaacaaaatgaacgaataa 1896
1905 GATATTGTGAATTTACACATACAAAAAGTTTATTATTAATAAAAAATTAAGACATTAA 1964
1897 aagactcaatgtattttaaanaatgtagaatlaaataataatataatgttcccaattct 1956
1965 AAGATAGAAATTTATTTTAAATAATGCACAACTAAAAAGATGATTCATACCAATATT 2024
1957 tacaacaagaataaagaagaacctaattatattgaatgttgaagaaagaatgtat 2016
2025 TATTAACCAAAATAAACCAGAACCATTTTATTATTGTATTAAAAAAGAGTAGAT 2084
2017 aaattaaagttgtatgtcctaagtagaatcatctgaataatgaagaaaaaatacata 2076
2085 AATTTAAAGATTTTATTCACAAAGTAAGACATGTTAAAGAAACACAACTGTCTTA 2144
2077 aaacagaaggtcaatcgtgaataatcgtgaacacatcaaccgaaggaataaagaagaca 2136
2145 TCAAGATATACCAACCTTATGTTAGTGACAGCAAGCAAC---AAGTGAAGATGGGGTCA 2201
2137 gcaactacaacacctcggacaagaagatctgtcttgaaggaagatcagtagacaagca 2196
2202 TCCACACACACATTTATCCCAATCAGAGAAACAGAAACAGAAACAGATAGACAC 2261
2197 caagcaagaagaacaaaagaacacacacacagtagtaccagtagtaccagaagaaca 2256
2262 ACACACACGTTACATTAACATTAACACACACACACACACACACACACACACACACAC 2321
2257 gcaacagttcccaac 2316
2322 GTTGTGAAATTTCAATAGACATTAAGATTAAGACATTAACCAAGCCCTTGACAAAAA 2381
2317 gattactctgaataatataatgtattttaaactcatalatataatgtcaacataatct 2376
2382 GTTATCTTAAGAAATTTGATGAATTTTAACTAAATCATATATGTCATTAATATT 2441
2377 ttggtttcacactcaactaagaagaatataaagaatataaataatgaagaag 2436
2442 TTAGATCAAACTGTAGTAGGACCAAAATTAATTAAGAGTATTAATCTTCTCCAGAA 2501
2437 gaagaagaacaaatlaagttcagtagtccatagacttatgtttaataatacaaaataac 2496
2502 GAAAAAATTAAGATTAATATCATGTGATTCATTATTTTAAATTCAAAAATAAC 2561
2497 ataactgttaatgtatctatgttgaatagcttaacaataagtttatcacactattatg 2556
2562 ATACTGCTATGATTAAT 2621
2557 gaaatttgaagaagaatgtgtgtaattatataataaataaagaatataagaacaaatt 2616
2622 GATTTATATCAAAAGAAATGATTTATTTATTAATCAATAAAGAGAAAGAAATTCATC 2681
2617 aaaaattataggaagcgaaaaaa-----gtatccacatctgttaaaactcttca 2670
2682 AAAAAATTTATGAGGAGCAAAACAAATTAATCAATCAATCAATCAATCAATCAATCA 2741
2671 agtcatcaatgtcaacatataatcaataacactcaagataaaccga----- 2717
2742 AATACAAACGTTAAATATGCTCAATCCGCAACTCACAGTAATTTCCCAAAACCAACATCA 2801

2718 -----agta 2721
2802 AATGATCTCTACCAATACCCAAATGGTGTAGCTGATCATCTGCTGCTGAGTT 2861
2722 agtgcacaatgtatatacatcacaatcttacaatattggaataatagtttaaatatttgaa 2781
2862 GAAGAAAGCATGATGCCCTTAACAGTATTTGCTAATTAAGCAATTTGAAGTATTGTT 2921
2782 aacatattgaatcttggaaaaaacaataatataacaaagttt---atagtgcaaaa 2838
2922 AGTCTTTAATCTTGGAAATTAACCTTAAGTACCTTAATCCATTACCTTTCTACACA 2981
2839 agtagtgaacactttatataaagaatataaagaatgtatataatattataatgaatct 2898
2982 GAGATGGAATAATTTATGAGATATTTTAAATAATTAATGATATTTTATGATGAT 3041
2899 ttacaacattctgtaaatctaaagctgaatgaatlaatcaatctgaatgaatgaatcaaa 2958
3042 ATCAAAACATTTGTAAATCTAATTCAAAAGTAATTCACAGTTTGA---CCGAACACAA 3098
2959 aggaagaattgaagaagataatataaataaataaacttcaagttatcatlgtat 3018
3099 AAAAAATGCAATTAATGATGAATTAATAAATTAAGATCACTTACAGTTATCATTTGAT 3158
3019 ttaataataataataataatagaagaatattgataaagaagaagaagttgt 3078
3159 TTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3218
3079 aaataaataatgaataataaataacttacttataaagaacaaatagaatcaaatgt 3138
3219 CAAGACAAATGCAATTAATAAATACTTACTTTATTAAGAAACAAATTAAGCAAAATTTG 3278
3139 aattcaacttaataacccaagcatgtatataaacttttctgttttcttatacaaaaaa 3198
3279 AATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3338
3199 aaagaagctgaatagcagaactgaacacacatlaggaacacacacacacacacacacac 3258
3339 AAGAAAGCTGAATTAACACAACTGAATAACACATTAAGAAACACAAATTAATTTGAAA 3398
3259 catataaagaagctgttataatataatattgtagaatactccatataaacttaagt 3318
3399 CATTATTAAGGACTGCTGTAATATTATTAATGATGATCAATTCATTAATAAATTTAAGT 3458
3319 gaagaatcaatccaaagaataatattgccagtttgaagaacttgaagtataagt 3378
3459 GAAGTATCAATTAACAAACAGATTAATTTGCCAATTTGAAAAATTTGAGTATTAACT 3518
3379 aatttagaagaagaatlaaagataatlaaatttagaagaagaagaataatcatcatact 3438
3519 AAAAAATGATGAATACTCAATGATTAATTTACATTTTAGGAAGAAATAATTAATTTCTTA 3578
3439 tcaagtgtatlaacatcatltaatttgcgtgaattlaaagaagtaataaataaataatatt 3498
3579 TCAAGTGAATTCATCATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3638
3499 acagtgtaattctccaaagtgaaataatatacagtagtataaataatgaatcttacaaga 3558
3639 ACAGGTAATTTCCAAAGTGAAATTAATTAAGAAAGTTAACGAAGCTTTAAATCTTAAGCA 3698
3559 aaatttccccaagaagaacagat----- 3582
3699 AATTTTCTCCGAGAACCAAAAGTTTACACAGTTGTAACCTCCACCTTCACACATGTAAT 3758
3583 ---gttcaacagttgtaagtgaaagtgatccgacacatlagaaacaaagtgcaacaaag 3639
3759 CCATGTCATTTATGCTAAGGAGTAGTGTAGTTCCAGATTCACAAAGAAAGAAACACAA 3818
3640 aaacagaatcaacatcagtaggaagcagagcttaacacaaataacacacacacacacac 3699
3819 ATACCAACTTTCAGGCTCTTTATTAATTAACAGAAATTAACAAAGATGATCAACAAATTA 3878
3700 gatgtagaagtagatgagtagtataatcatataattggaagaatccgaagaagattat 3759


```
ENEVDYIKPLAGVYRSIKKOIEKNITFENLNDILNSRLKKRRYFLDVLSDLMQ
FKHISNEYIIEDSFKLNSDEQNTFLKSKYKIKSEVENDIKFAOEGISYKYLAKY
KODLESIKVIEKEKEKPEPSPPTPKADDEQKESKFLPLNIETLVNINLVNK
IDDYILINKAKINDCNVEDFAHVKITTKSLDIAIDDKIDLEKFNDEFAIKKLIIND
TKKMLGKLSLSTGLVONFNPNTISKLIBEKPDMINISOHGVKKOCPENSCFRIID
EREKCKLINTKOEGBKVENPNPCNENNGCGDADATCTEEDSGSRKKTICECTKP
DSYPLFDIGFICSSNFIISFLIIMLILYSEI"
193. .5238
mat_peptide
/product="surface antigen (AA 1-1682)"
463. .465
misc_feature
/feature="pot.glycosylation site"
850. .852
misc_feature
/feature="pot.glycosylation site"
1543. .1545
misc_feature
/feature="pot.glycosylation site"
1741. .1743
misc_feature
/feature="pot.glycosylation site"
1954. .1956
misc_feature
/feature="pot.glycosylation site"
2539. .2541
misc_feature
/feature="pot.glycosylation site"
2575. .2577
misc_feature
/feature="pot.glycosylation site"
2830. .2832
misc_feature
/feature="pot.glycosylation site"
2880. .2882
misc_feature
/feature="pot.glycosylation site"
2890. .2892
misc_feature
/feature="pot.glycosylation site"
3028. .3030
misc_feature
/feature="pot.glycosylation site"
3106. .3108
misc_feature
/feature="pot.glycosylation site"
3400. .3402
misc_feature
/feature="pot.glycosylation site"
3733. .3735
misc_feature
/feature="pot.glycosylation site"
4897. .4899
misc_feature
/feature="pot.glycosylation site"
5032. .5034
misc_feature
/feature="pot.glycosylation site"
BASE COUNT 2376 a 685 c 702 g 1629 t
ORIGIN
Query Match 55.58: Score 2732; DB 96: Length 5392;
Best Local Similarity 73.98: Pred. No. 2.3e-245;
Matches 3795; Conservative 0; Mismatches 1095; Indels 246; Gaps 16;
1 atgaagatcatactcttctatgctcatctcttcttattataataacaaatgta 60
136 ATGAAGATCATATCTCTTTTATGTTATCTTTTATTATTAATACACATGTGTA 195
61 acacatgaagatlatcaagaactgtcaaaaactagaagctttagaagaatgcagatg 120
196 ACACATGAAGATTATCAACAAGCTTCAAAAACTGAGAGCTTTAGACAGTCAATG 255
121 aaagttatagttatttcaaaaagaaatggtatcaatgaagaagaacagtgaaaca 180
256 AAGAGTTATAGTTATTCAAAAAGCAAAATGATTAATGAAGAACCAATGGAACA 315
181 gctgttaacaactagtcacactgttcaaaaggttcagtcctcagtgagttcagtg 240
316 GCTGTTAACAACAGTACACTGTTCAAGTGTCAAGTACTTCAAGTGTG----- 365
241 tcaagttgctcagtgagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 300
366 -----TTTCAGTTCCTTCAAGTGTCTTCAAGTGTCTTCAAGTGTCTTCAAGT 417
301 gcttcagtggttcagtgatccaagcgttacaagccttcagatataatccaagtgat 360
418 GCTTCAGTGTGTTCAAGTGTCAAGCTTCAAGTGTCTTCAAGTGTCAAGTGTCA 477
```

```
361 gacttaaaactctacgcgtgatttaaaacacagagtaacgaataacttgtaacatacaaa 420
478 AATAGTAAGAACTTACCGTGTATTAACAAATAGAGTTCATAATTTACTGTTACTATTA 537
421 gaactcaaatatacccaactccttgatttaactaataatgtaacttggtgtaata 480
538 GAACCTAATATCCCAACACTTGTGATTAACCAATCAATATGTTGACTTATCTAAATA 597
481 atcaatggtttcaaatatttaattgataagataagaagaataatgaataatataaa 540
598 GTTGAATGTTTCAAAATATTAATTTGATGATGAAGAAATTAATGATTAATTAATA 657
541 ttaactttatttgatttatttaagaagcaaaataaagatgataagtcataatgat 600
658 TTAACTTTTATTATGATTTATTATTAAGACCAATTAATGATGATGATGATGATGAT 717
601 tgtcaaatcccttcaactccttaaaatcgtgcaaatgaattagaagtaacttaaaaa 660
718 TGTCAAAATACCTTTCAATCTTAATAATTCGTCAAAATGAATTAACGACTTAATAA 777
661 gtttgcgatatagaagaacatagaacataataagaataatgaagaagaat 720
778 GTGTTGGATATGAAGAACCATTAAGCAATTTAAAGATTAATGAGAAATGAAGAT 837
721 tacattaaaaaaataaaacacacacacacacacacacacacacacacacacacac 780
838 TACATTTAAAAAATATTAACACACATGCAATATTAATGATTAATTTAGAGAAATGA 897
781 aaaaacttgatataaaataagaatgcaactaagaagaagaagaagaagaagaagaaga 840
898 AAACAATGATGATCAAAATTAAGATGCAATGATGAAGAGAGAGAGAGAGAGAGAG 957
841 gttcaatgatctctctctatcaataaacaatagaagaagacacataatcaataag 900
958 GTCATATATATATCTTTTATTTTAAATTAACATTAACATTAACATTAATTAATTA 1017
901 gtttgaagaacgatactgacacttcaaaaagaagaagaagaagaagaagaagaaga 960
1018 GTTTTGAAAAACGATTTGACACTTTAAAAAATGAAACCTTAAGAAATTAATCTTGA 1077
961 aagataaatgaataaaa--atccccaacgcgcgaactccttgaaatatacacaata 1017
1078 GATATAGATTAATTAATTAACAGATGCGCAAAACCCACTACTGAGATTAACCAATTC 1137
1018 ctctctg---ataagaacaaaaaatcgaagacacgaagaagaagaagaagaatctgc 1074
1138 CTCCCTGAGAAATTAAGAAAAAGAGTCGAGGACACGAGAGAGAGAGAGAGAGAG 1197
1075 aaacatttaaatataatgataagttatctactgacactgaactgaactgaatact 1134
1198 AATATATTAATTTATACATTTGATTTATTTACTGATCCACTTGAATTAATTAAT 1257
1135 ttaagaagaagaagaataaataatataatgaatgaagaagaagaagaagaagaaga 1185
1258 TTAAAGAGAAAAAATTAAGAAAGTTGATGTAACACCTTAATTAACAGATCTCAAG 1317
1186 gaatacactgaacccaagaataatccaagaatgaacttaactccttgatataacg 1245
1318 GTTCAATATCAAAAGTCTTATTCCTTAATGATGATGATGATGATGATGATGATG 1377
1246 ataaactgaactta-----aagaaactaaactccttggtgataatataatca 1296
1378 ATTCAATATTCATTAACCTGCAAGTATGATTAATAATTTCAATGATGATTAATGA 1437
1297 ttgattatacaaaagaacaaaglaaaacatalatactgaataagaagaagaagaaga 1356
1438 GATACATAAGAAAAAATTAATGAAAAAATTAATTAACAGATTAATTAAGAGAAAA 1497
1357 ataaatgaataatgaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1416
1498 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1557
1417 tcttaagaagaagactaagtcctttaaattataacaaagaagaagaagaagaagaat 1476
```


D	1558	CAAAATTAATAAATTTACTTGACATTTATGAAAGTCAAAAAGCATTTATGAGAAATTTACTT	1617
O	1477	aaatgaattatgatgatgcaaatccaaataaataatagatttaacttaacttaactgaaataatg	1536
D	1618	GAAAAATTTTATGAAATGAAATTTATATATATTTTGACAAAGATGTCGTATGATTAATAA	1677
O	1537	atgggttaaaagatatcatataaagtltgaaataactacacacataacttctgacac	1596
D	1678	TTTCAGTGCAGATATATACATATATATGTTGAAAAACAAAGATATATATATATATATATAT	1737
O	1597	tatgaaatcttcaaacatactctgaaaglttaacaaagctcttaataatataatgaaagat	1656
D	1738	TCTATATATCTTGATATATATGTTTCAAAATTTAAAAAGGCTCTTTCATATCTTGAAGAT	1797
O	1657	tacttctaaggaaataatgltgaaaaaagaaatlaaaatcatataaaatlttaataagc	1716
D	1798	TATCTTTTAAAGAAAGAAATTTCTGAAAAAGATTTTAATCATTTATTTACTTTGAAAACT	1857
O	1717	aaaataagaataatgagatctgaaacataatgltgaaataatlaaaaaagatgaaagacgtc	1776
D	1858	GGCCTTCGAAAGCTGATATATAAAATTTAACGAAGAAATTAAGATGAGTGTGAAAAACAAAT	1917
O	1777	tttggaaaaaaaatacttaagacgaaataaaccgagatgaaaaaatlttaagaatctc	1836
D	1918	CTAGAAAAAAA---TTTAAAGGACTTACACATTCAGCAAAATGCTTCTCTTGAAGATCT	1974
O	1837	gacatgtataaagtacaaglttcaaaaaglttlaataatgaaacaaatltgacgaataa	1896
D	1975	GATATTTGTAAATTTACACATACAAAAAGTTTATTTATTTAAAAAATTAAGAAAGCTTATAGA	2034
O	1897	aagactcaatctgactttaaanaaagttagaataatlaaacataatatacatgttcccaatctc	1956
D	2035	AAGATAGAAATTTTATTTAAAAAATTCACAACTGAAAGATGATATGATGATACCAAAATATT	2094
O	1957	tacaaacaaagaaataagcaagaactcttatataatgtgtgtgtaaaaaaagaatctgat	2016
D	2095	TATTAACCAACAAATTAACACGAACCATATTTATTTATTTAAAAAAGAAAGATAGAT	2154
O	2017	aaatlaaaaglttctgactgaagtagaatctatgaataatgaaagaaaaaataaata	2076
D	2155	AAATTTAAAGAAATTTTACCACAAAGTAAAGACATGTGTTAAAGAAAGACAAAGCTGTCTTA	2214
O	2077	-----aaaacgaaggtlcaatctgataatctcgaaacccatcaaccgaagagaataaca	2130
D	2215	TCAAGATTTACACAACTTTAGTTGCACGAAGGAAACACTGAACATGGGGGCTACCTCC	2274
O	2131	ggacaaagcaactacaaaactctgacaaacgaagcagatctgctttagaaggaatctcgtat	2190
D	2275	ACACACACATTTATCCCAATTCAGGAAACGAAGTAAACAAAGAAACAGAAAGTAAACAGAA	2334
O	2191	caagcaagcaagcaagaataacaaacgaacaaacccacagctaccagctacacagctacgaag	2250
D	2335	GAAACAGTGTGGACACCAACACACGGTAAACATTTACCTTACCACCAAAAGAAAGATCAGCA	2394
O	2251	gcaaaagcacaagctcccaacacacacgaagcaccagtaataatlaaaactgaaat-----	2304
D	2395	CCAAAGAAAGATTAAGTTGTTGAAATTTCAATTTGAACATTAAGATGATGACATTTCCAA	2454
O	2305	-----gttcccaatttagatctctgaaaaatataatgatcttcttaactctcatatata	2361
D	2455	GCCTTGACAAAAACAGTTTATCTTAAAGAAATTTAGATGAATTTTAACTAATATATATATA	2514
O	2362	ttctcaaaatataitltgttctcacactcaactatgaagaagaagaatataaacaatat	2421
D	2515	TGTCAATAAATATATTTTATCTCAAACTCTAGATGAGACCAAAATTTATTTAGGATATAT	2574
O	2422	aaaattacaagaaggaagaagaacaaatlaagltcagctcatcatcagaactatgctt	2481
D	2575	AATCTTACCTCCAGAAAGAAAGAAAGATTTAAATCATGTGATCATCTAGATTTATATTT	2634
O	2482	aatatataaataacatactcgttaatgtatctlaigtltgatagtcttaaacaaatagttat	2541
D	2635	AAATTTCAAAATTAACATCTGCTATGTATTTCTATATATGATAGTATGACAAATGATTTA	2694
O	2542	tcacaactattatgaaattatataaaagaanaatgglttgtaattatataaacttaag	2601
D	2695	CAACATCTCTTTTGAATTTATATCAAAAGAAAGATTTTATTTATTTACATTAACCTAAA	2754
O	2602	gataatgcaaaatataaatttatatgagaagcgaaaaa-----gtatccaactc	2655
D	2755	GAGGAAATATACATCAAAAAATTTATTTAGAGGAGCAAAACAAATTAACCTGCAACATCT	2814
O	2656	gtaaaaactcttcaagctcaatcgaatcgaacacatcatcaaacaccatgaataaaccc	2715
D	2815	ACATCCAGTCTCGAATTTACACCGTAAATTAAGTCTCAATCCGAACTCACAGTAAATCC	2874
O	2716	ga-----	2717
D	2875	CAAAACCAACATTAATGATCTCTTACCAATTAACCAAAATGCTGATGATCACT	2934
O	2718	-----agtaagctgcaaatgataatgatacatcacaatctcaaatltgaaataatg	2766
D	2935	GGTCTGCTGTAGTTGAAGAAAGCATGATCCCTTAAACAGTATTTGCTATTTAGTAACGAT	2994
O	2767	ttaaatcttctgaaacataatgtagctctgaaanaaacaanaatatataccaagattat	2826
D	2995	TTGAAAGCTATTTGTTACTCTTTTAATCTTTGGAATTTAAACTTAAAGTACCTAATCCATTA	3054
O	2827	---ataggtcaaaaagatgltgaaactcttatgaaagatatlaaagaatagatgataca	2883
D	3055	ACCATTTTCTAACACAGAGATGGAAGAAATTTTATGAGATATTTTAAAAAATTAATGATACC	3114
O	2884	tttataatgtaactcttccaactcttgtaaaacttaagctgataatataatctcatc	2943
D	3115	TATTTATATGATGATATCAAAACAAATTTGTAATTTCAAAAGTAATTAACGATTTG	3174
O	2944	aatgataatcaaaaggaagaaatltgaagaaatataatlaaaataaanaaaactta	3003
D	3175	A---CCGAAACACAAAAATTAAGCTTAATTAATGATGAATTTAAAAATTTAAAGATACCTTA	3221
O	3004	cagtlcatcatctgattatataataataataatlaaaatlaagaagatlaattgataaa	3063
D	3232	CAGTTATCATTTGATTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAG	3291
O	3064	aagaaacaglttglttaataataatgcaaatlaaaaaacttacttataaagaacaa	3123
D	3292	AAAAAAGAACTTGCGCCAAAGCAAAATTCAAATTTAAAAAAGCTTATTTTAAAAAGAAACAA	3351
O	3124	ttagaatcaaatgtaattcaacttaataaaccgaagcagatgataacaaactcttcgt	3183
D	3352	TTAGAAATCAAAATTTGAATTTCACTTAATTAACCCATATATGATTAATTAACAAATTTTGT	3411
O	3184	ttcttacaacaaaaaagaagcgtgaatatagcagaactgaaacacacatagaanaacaa	3243
D	3412	TTCTTTTACAAAAAAGAAAGCTGAATATGCAAAACTGAAACACATTTTGAAGAAACACA	3471
O	3244	aaaatattatgaaacatlaaagaagcgtgttaaatatataatlaaagtgtgaatctcaca	3303
D	3472	AAATATTTATTTGAACATTTTAAAGGACTTGTAAATTTATATATGATGATGATCAATCTCA	3531
O	3304	ttaaaaacttaaggtgaagaataatccaaacgaagaataatctgcagtttgaanaac	3363
D	3532	TTTAAAACTTTAAGTGAAGATCAATTTCAAAACAAAGATATTTATTTGCAATTTTGAAGAA	3591
O	3364	tttaagctatctgaatgaaga	

QY	1657	tatctttaaaggaatagtagtltgaaaaaaagaaattaaatataaataattaaagc	1716
Db	1783	TATCTTTTAAAGAAAGAAATTTCTGAAAAAGTTTTAAATCAATTTTAACTTGAAACT	1842
QY	1717	aaaaaagaagaagatgaacactagttggaataattaaaaaagatggaagacagctt	1776
Db	1843	GGCCTCGAACCTATATTAATAAAATTTACAGAGAAATTAACAGTAGTGAACAAATTT	1902
QY	1777	tttgaaaaaaaatlaactaagaacggaanaaataaacagatgaaaaaatttgaagatct	1836
Db	1903	CTAGAAAAA---TTTAAAGACATACACATTTGACGAAATGCTCTTGAAGTACT	1959
QY	1837	gaactgttaaagtaccaagtlcaaaaaagtttacttaatgaacaaaatgtacognatlaaa	1896
Db	1960	GATTTTGTAAATTTACAGTACAAAAGCTTTTATTAATTAATAAAATGTGAAGACTTAGA	2019
QY	1697	aagactcaatgtgtttaaaaaatgtagaataaacaataatacatgttcccaattct	1956
Db	2020	AAGTTAGATTTATTTTAAAAATGCAACACTTAAAGATGATTTCACTGACCAATTTT	2079
QY	1957	tacacaagaagaaataagaacgaaccttacttaattgtgttgaaaaaaagattgat	2016
Db	2080	TATTAACACCAAAATATTAACGAGAACCATTTATTTTAATGTATTAAAAAAGAGTACAT	2139
QY	2017	aaattaaagtgltcagtcactaagtlagaacatcattgataaaltgaagaaaaaaacata	2076
Db	2140	AAATTAATAAAATTTTATACCAAAAGCTTAAAGACATGTTTAAAGAAACAAAGCTGTCTTA	2199
QY	2077	-----aaacagaagtgtaactcgtgtaatctggaacccaataccgaaggagaataaca	2130
Db	2200	TCAAGTATTAACACACCTTTTGTGACGACAGCAACCACTGAAGATGGGTGCATCC	2259
QY	2131	ggacaagcaactcaaaacctgtgcacacagaagagatcgtcgttgaaggagatcagta	2190
Db	2260	ACACACACATATTCCACATCAGAGAAACAGAAAGTATACGAGAAACAGAGTAAACAA	2319
QY	2191	caagcacagaacagaacaaaaaagaacacaaaccaacagtaacagtaaccagtlaccagaa	2250
Db	2320	GAACACATAGCACACCAACACGCTTAACAAATTAACATTACCCAAAGAAAGAAATCAGCA	2379
QY	2251	gcaaaagcagaagtcaccaacacacacagacagcaagtaataataaacctgaaat-----	2304
Db	2380	CCAAAGAACTTAAAGTGTGTAATAATTAAGAACATTAACAGTAAGTAAACAAATTCACA	2439
QY	2305	---gttccaaattagatlaactcttgaaaaaatataagaatttttaaatlaactcatata	2361
Db	2440	GCCTGACAAAAACAGTTTATCTCAAAAGAAATTAGATGAATTTTAACTAAATCATATTA	2499
QY	2362	tgtaacaatatalttgglttcacactcaactatgaacgaagaagataltaaaacaatat	2421
Db	2500	TGTCATTAATATATTTTAGATGCAAACTCTAGTATGACCAAAATTTATTTAGCGTATAT	2559
QY	2422	aaaaattacaagaagagaagaagaacaaatlaagtlcatatgataccacttgactattgtt	2481
Db	2560	AATTTTATCTCCAAACAAAGAAATGAAATTAATTAATCATGTGATCCATTGTGATTTTATTT	2619
QY	2482	aataataaataacacactcgtgaagtaacttcatatggttgaagttgaacttaacaatagttta	2541
Db	2620	AATATTCAAAATTAACATACCTGCTATGATTTATTAATATATGATAGTATGAACATTGATTTA	2679
QY	2542	tcacaactattatggaatattatgaaaaagaagatggtltgtaattatataaacttaag	2601
Db	2680	CAACATCTCTTTTGAATATATCAAAAGGAATGATTTATTTATTTACATTAACATTA	2739
QY	2602	gataatgacaaaattaaatatattatagagggaacgaaaaa-----gtatcaacactc	2655
Db	2740	GAGGAAATATACATCAAAAATTTATTTAGGGAGCAAAAACAATTAACGTGAACATCACT	2799
QY	2656	gtaaaaacttttcaagtlcatatgaatgcaacattatcatatgaacactcagatataacc	2715
Db	2800	ACATTCACGTCTCTGGAATATACACCGTAAATTACTGCTCAATCCGCACTACACAGTAATTC	2859
QY	2716	ga-----	2717

Db	2860	CAAAACCAACATCAAAATGCATCTCTACCAATACCCAAATAATGGTGTACGTATCAATCT	2919
Oy	2718	-----agtaagtcgaatgtagatcacatctctacaatttgatatagat	2766
Db	2920	GGTCTCTGCTAGTTGAAGAAAGTCATGATCCTTAAAGCATTTGTGTATTAGTAAAGAT	2979
Oy	2767	ttaaatatlttgaaacaacatattgagctcttgaaaaaacaataatataacagaatta	2826
Db	2960	TTGAAAGGATATTGTAGTCTTAAATCTTGGAAAAATTAAGTACCTAAATCAATTA	3039
Oy	2827	---tagtgcaaaaagtagtgaaacttttagaaaagatitaaagatagtagaca	2883
Db	3040	ACCAATTTCTACACGACGAGGAAAAATTTATGACAAATTTTAAATAATATATATCC	3099
Oy	2884	tttataatgaatcitttacaanaatttcgttaaaatctcaaaagcgtatgatlaatcatctg	2943
Db	3100	TATTTTATGATGATATCAAAACAATTGCTAAATCTAATTCAAAAGTATTACAGTTTG	3159
Oy	2944	aatgaatgaatcaaaaggaagaaatagaagaagatatataatataaaaaactta	3003
Db	3160	A--CCGAAACACAAAAAATGCAATTAAGATGAATTAATAAAATTTAAACATCTTTA	3216
Oy	3004	cagttacatcttgattatataataatatataatataaattgaagaattatcttgataa	3065
Db	3217	CAGTTATCATTTGATTTATATATTAATTAATTAATTTAAATTGATATGATTTATTAATAG	3276
Oy	3064	aagaaaaaglttgtagaataataaatgcaattataaaaacttacttataaagaaca	3123
Db	3277	AAAAAGAACTTGCGCAGCAAAAATGCAAAATTAATAAACTTATTTAAAGAACAA	3356
Oy	3124	ttgaatcaaaatttgatttacttaataacccaagcatgatatcaaaaacttctgt	3183
Db	3337	TTGAATATCAAAATTTGAATTCATTAATTAACCCATTAATGATATACAAAATCTTTCTGT	3396
Oy	3184	ttctttaaacaanaaagaagcctgaaatagcgaagaacgtgaacaacatagaanaaca	3243
Db	3397	TTCTTTAAACAAAAAAGAAGCTGAATATGCGAAGAACTGGAAACACATTAAGAAACAA	3456
Oy	3244	aaaatcatltgaaacatitaaagaactgttaaatatataatagtgtaacatctca	3303
Db	3457	AAATATATATGAAACATTTATTAAGACCTGTGTAATATTTAATAAGTGCAATCATCTCCA	3516
Oy	3304	ttaaaaacttaagtagaagatcaattcaaaaggaagatlaatagccaaglttgaanaac	3365
Db	3517	TTTAAAACTTTAACTGAGATATCAATTTCAACGAAAGATTAATTAAGCCAAATTAAGAAAA	3576
Oy	3364	tttaagatlatlaaglaaatagaaggaanaattlaaagatlaattlaaattlaaagaana	3423
Db	3577	TTTGAAGCATTAAGTATAATATAGATGGAAGAACTCAATGATTAATTTACATTTACGAAGAAA	3636
Oy	3424	aaatcatcatctatacaagtggaatcacatcaattlaattgtcgtgaatlaaagaatlaa	3483
Db	3637	AAATATATCTTTCTTATCACTGATGATTAATCATATTTAATTACTGCAATTTAAACAACTATA	3696
Oy	3484	aaaaataaaaattatacaagtaattctcaaaaggaataatlaaagaatgtagaagtagca	3543
Db	3657	AAAAATTAATAATTATACAGTAAATTTCTCCAAAGGAAGAAATTAATTAAGAAAGTTTAACGAACT	3736
Oy	3544	ttgaatcitttacaanaaattctcccaagaaggaacgat-----	3582
Db	3757	TTTAAATCTTACGAATAATTTTCTCCGAAAGCAAAAGTTTACAAAGTTGTATCTCACACT	3816
Oy	3583	-----gttcgaacagltgtgaagtgaagtgaatgcgacacatagaa	3624
Db	3817	CAACAGATGTAATCTCATCTCCATTTATCTGTAAAGGATGATGATGTTCCAGATGCCACA	3876
Oy	3625	caaatgtaacaaagaacacagcatcaactcaltgtagagcgaggtcttaacaaatlaa	3684
Db	3877	AAAGAAAGAAACAAATTAACCAACTTACAGGCTCTTTATTTAACGAATTTACAAACAGTAATA	3936
Oy	3685	acatcacaaatgtcgaatgtagaagtagaagatcaatcaatagtaactatatttggagaa	3744

D	b	3937	CAATTACCAAAATTTTGACGGAAGAAGATGATCTCTTGTAGTTGTATATACCACTTTTGGACAA	3996
O	y	3745	tcggaagaagaattatgtaattgaagcaagtatgatacaggaagacgaatacctctcc	3804
D	b	3997	TCCGAGCATATATGACGAAATATTATGATCAAGTGTAGTAACGTGAGAGCAATA---TCTGTC	4053
O	y	3805	gtaattgatacactcttctaataattgaaatgaatgatgaggtttattattaacct	3864
D	b	4054	ACAAATGATATATTCCTTCAGGATTTGGAATAAGATATGATATATATTTAAACCT	4113
O	y	3865	ttgaagatgtctatagaagtttaaaaaaacatgtagaataaagctatgaactttat	3924
D	b	4114	TTAGCTGAGATATATACAGCTTTAAAAACCAATTGAAAAAACCTTTTACATTTAT	4173
O	y	3925	gtlaatgtlaaggaatttttaaatlccagatlttaataaagctglaaattlccaaatgt	3984
D	b	4174	TTAAATTTGAACGATATCTTAAATATTCGCTTAAAGAACAAATTTTCTTACATGTA	4233
O	y	3965	ttgaatcgaatttcaattccatatagaattacaatcaagtaattatgtctcaagat	4044
D	b	4234	TTGAAATCTGATTTTAAAGCAATTTAAACATATTCCTCAAAAGAAATACATTATTGAAGAT	4293
O	y	4045	ccataaattctctataaagaaaaaggaataaattcttaagcaattataatatt	4104
D	b	4294	TCAATTTAATATTTATGAAATTCAGAACAAAAACACTTTTAAAAAGTTACAAATATATA	4353
O	y	4105	aaggattcctaagaatacgaatataaaatttcgcgaatgactctcgtgatataataa	4164
D	b	4354	AAAGATCTAGTAGAAATATGATTTAAATTTGCACAGGAAGATATAGTTATATGAAAG	4413
O	y	4155	ttatccgaaaaataataaaccgaattgagattcaattaaaaatataaac-----	4215
D	b	4414	GTTTATGAGAAATATTAAGATGATTTAGATCAATTTAAAAAGTATATCAAAAGCAAAAG	4473
O	y	4216	-----gacaaaac	4224
D	b	4414	GAGAAATTCATATATACCAACACACACACCTCGCTCACGCAAAAAACAGCACACAA	4533
O	y	4225	ggtgaanaatggaatactcctcccttttaaaacaatatgagacctatataaacgct	4284
D	b	4534	AAGAAGGAAGATAGTCTCTTCCATTTTAAACAACTTGTAGACCTTTATCAATTAAGCTTA	4593
O	y	4285	aatgataaattgatttattctgtaattcatattagaagcaaaagctctaaatatacata	4344
D	b	4594	GTTTATTAATAATGTGCGATTCCTTAATTAATCTTAAAGCAAAAGATTAAACGATTTGTAATGT	4653
O	y	4345	gagaaatcgaacgtagagagtttaaaagaacttaacttacttaaaacaatcgaac	4404
D	b	4654	GAAAAAGATGAGACACTGTTTAAATTAATTAACCTTAGATTTTAAAGCAATTTGATGAT	4713
O	y	4405	aaattgacgaattttaaanaaaataaacaattcgttgaatgctcgattatcaacagat	4464
D	b	4714	AAATATAGATCTTTTAAAAACCTAAGCACTTGAGAGCAATTAAAAATTTGATAATGAT	4773
O	y	4465	tataacataataacttattgacaaagttccctagtagaagtaiggttttggaaactt	4524
D	b	4774	GATTCGAAAAAAGATATGCTTTGGCAAAATTAATCTTAGTACAGGATTTAG--TTCAAAAATTTT	4830
O	y	4525	gctaaacgcttattatcatttacttctgttgaactgcgaagtgatgtttaaactttca	4584
D	b	4831	CTTATATACAAATATATCAAAATTAATTTGAAGGAAATTCACAGATATGTTTAAACTTTTCA	4890
O	y	4585	caacacaaatgctgtanaaaaaaacatgcccacaattctcgaatgtttcaagacattagat	4644
D	b	4891	CAACACCAATCGTAAAAAACAATGCCCCAAGAAATTTTCGATGTTTTCAGACATTTAGAT	4950
O	y	4645	gaagagaagaatgtlaaagtlttataaattacaacaagaagtgataaattgtltgaa	4704
D	b	4951	GAAAGACAGATGTAAATGTTTATTAAATTAACAACAGAGGTGATTAATATGTTTGA	5010
O	y	4705	aatccaatacctactgttaacgaaaataatggtgtgtgtatgcagaatgcacaatgtacc	4764
D	b	5011	AATCCAATCTTACTTTGTAACGAAATATATGTGGATGTGATGACAGATGCCACATGTAGCC	5070

QY	4765	gaagaagatcagctgagcagcaggaagaataatcacaatcgaatgtaatgtaactaaacccgatctt	4824
Db	5071	GAAGAAGATTCAGGTACGACGAGAAAAAGAAAACACATGTGAATGTAAACCTAATTC	5130
QY	4825	tatccacttccagatggaattcttcgcagcttcctcctaactcttaggaatcatctcta	4884
Db	5131	TATCCACTTTTCGATGCGTATTTCTTCGCAATTCCTTAACCTCTTGAAGAAATATCATTTCTTA	5190
QY	4885	ttaataactcatgtaataatatacatagttcatctaa	4920
Db	5191	TTAATACCTCATGTATATATTATACAGTTTCATTTAA	5226
RESULT	14		
PEMEZSA1B			
LOCUS	5312 bp	DNA	INV 10-MAR-2001
DEFINITION	P.falciparum gpi190 (MSA1, MSP1, PMMSA)	for precursor of major	
ACCESSION	235327	U09209	
VERSION	235327.1	GI:929795	
KEYWORDS	gpi190 gene; merozoite surface antigen; MSA1 gene; MSP1 gene; PMMSA gene.		
SOURCE	Malaria parasite P. falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 5312)		
TITLE	Pan, W., Tolle, R. and Bujard, H.		
JOURNAL	A direct and rapid sequencing strategy for the Plasmodium		
MEDLINE	falciparum antigen gene gpi190/MSA1		
REFERENCE	Mo. Biochem. Parasitol. 73 (1-2), 241-244 (1995)		
AUTHORS	96123395		
TITLE	2 (bases 1 to 5312)		
JOURNAL	Tolle, R., Bujard, H. and Cooper, J. A.		
MEDLINE	Plasmodium falciparum: variations within the C-terminal region of		
REFERENCE	merozoite surface antigen-1		
AUTHORS	Exp. Parasitol. 81 (1), 47-54 (1995)		
TITLE	95354793		
JOURNAL	3 (bases 1 to 5312)		
MEDLINE	Tolle, R.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (18-JUL-1994) Ralf Tolle, Abt. Prof. H. Bujard, Zentrum		
TITLE	fur Molekulare Biologie, (ZMBH), Univ. Heidelberg, Im Neuenheimer		
JOURNAL	Feld 282, Heidelberg, 69120, Germany		
COMMENT	On Aug 5, 1995 this sequence version replaced gi:535249.		
FEATURES	Location/Qualifiers		
SOURCE	1..5312		
	/organism="Plasmodium falciparum"		
	/isolate="NF54"		
	/db_xref="taxon:5833"		
	/chromosome="9"		
	/clone="3D7A"		
	/dev_stage="blood stage"		
	109..165		
sig_peptide	/gene="gpi190, MSA1, MSP1, PMMSA"		
gene	109..5271		
	/gene="gpi190, MSA1, MSP1, PMMSA"		
	109..5271		
CDS	/gene="gpi190, MSA1, MSP1, PMMSA"		
	/codon_start=1		
	/product="precursor of the major merozoite surface		
	antigens"		
	/protein_id="CAAB4556.1"		
	/db_xref="GI:929796"		
	/db_xref="SPRMEML:Q25922"		
	/translation="MKTIFFCSFLFIITINQCVTHESYOELVKLEALDAVLVGYGS		
	LFQEKWVLNLEEITTTGASGASGASGASGASGASGASGASGASGASGASGASGASG		
	SGPGTSPSSRSNTLPSSNTSSGASPADASDPAKSYADLKHSRVNYLFTIKLKP		
	ELPITLNMILTLDNINGFKYLIDYGEIEINLLYKLFNFDLIPAKINDVCANDYOD		
	PNFKITANEILDVLRKLYFGYGRKLRKINDVCKMEDYIKNNKTTININBLEGSK		
	TTIDKNADNEEGKRLYQAVDYSLYNKOLEEHNHLSVLEKRIIDLKNNENIRKLL		
	KLDEININPPANGSNPNLIDLNKRIIEEHEEKIKIANTIKFNIDSLFTDPLELEY		
	YLRKNNKVDVTPKSDOPTSVQIPKVPYNGIYPLPLDINHSLAADNDKNSGDL		

MPHTRKEKINEKIITDNKERKIFINNKKIKIDLEEKNIINHTKEQNKKLEDEYKESKRD
YELLELEKEFKEMNNPDKDNDYDKIFSAAYTNVVKORNNKFNSSNNVYVOKLAK
ALSYLEDESLRKISSEKDNHNYTLTKGLEADIKTEKESEKNNLEKNNKGLTHS
ANGSLSESDIVKLOVOKYLIKIDELRIELFELNAOIKSDHPNITKPPKPNKPPY
YLIYKKEVDKKEFIPKYKDKLKKKEQANLSSITOPPLVAASHTEDGGSTHTLSOG
ETEVEETEETEVEYGHVTTVTTPPTDPSPEKKEVENSIEKSNNSONLTKRTV
YIKRIDEPLTKSYIKHYILVSNSSMDOKLELVYNLTPEENELKSCDPLDIFNQN
NIPAMYSLDSSMNDLOHLFELLYOKEMLYLHKLEENHIEKLEQOIGTSTOS
SPGNTVTVAOSATSHSNOSQNSASTTONGVONDEBDSLYVLPVSHESHDLPVLSISND
LKGISVLNKGKTKVNPDLTSTTEMEKEFNIIKNDSTPNDNDIKOVKNSVYT
GTEETOKNANDEIKKIDTLOLSFDLYNKYIKLKDRLFNKKEIGDQMDOKIKLTL
KQDLSKLSLNNPNHVLONEVFPNKKKEALIEAETENLENTKILKHYKGLVRYN
GESSPLDKLSEVSIQTEDNYANLEKFRVLSKIDGLNNDLHLGKLSLSGLHLI
TELKEVINKNNTGNSPENNNKVNKALESYENLPKAVTVVTPPODVPSPPLSV
RYSGSSGSKETEOIPTSGLLTELVOOVONDEBDSLYVLPVSHESHDLPVLSISND
OYVGEALISVMDNLISGFENEYDVIYIKPLAGVRSJKOEIKENIFPNLNDILN
SLKRRKYFLVDESLDMOFRIISSNEYIIESFKLNSQNKNTLKSXKYIKESVEN
DIRFAOEGISTYEKVLAKDDESIRKVIKEKKEFPSSPTPPSPKATDEOKES
KFLPPLTNIEYLYNNLVNKIDYLIINLAKINDCNEVDEAHVKITKLSDLAIDKI
DLFKNPYDFEALIKLINDTKMDMLGKLGSTGLVONFPNTIISKLEGFODMLNISO
HOCVKOCPENSGCFRHLDERECKCLNLYKOEGRKCVENPPTCNENNGGCDADATC
TEEDSGSSRRKKTCECTPDYSPLFDGIFCSCSSNLFGLISFLILMLILXSF1"
BASE COUNT 2328 a 703 c 705 g 1576 t
ORIGIN

Query Match 55.2% Score 2713.8: DB 96: Length 5312:
Best Local Similarity 73.3% Pred No. 1.1e-243:
Matches 3791: Conservative 0: Mismatches 1117: Indels 267: Gaps 14:

OY 1 atgaagacataatcttcttaigtcatcttcttcttataataacacaaatgta 60
DB 109 ATGAAGATCATATCTTTTATGTCATTTCTTTTATATAATACACAAATGTGA 168
OY 61 acacatgaagaattataagaacttgcacaaactagaagctttagaagatgcagtatg 120
DB 169 ACACATGAAGAATTATCAAAAGCTTGTCAAAAGCTTTAGACAGTGCAGTATTG 228
OY 121 acaggtatagttatttcaaaaggaataatggtatataatgaagaagaatgtgaaca 180
DB 229 ACAGGTATAGTTTATTTTAAAAAGAAAATGGTATTAATGAAAGAAATTAATCA 288
OY 181 gctgtacacactagtaacactgttcaaaaggttcagttcaggtcaggtcaggtgc 240
DB 289 AAGGTGCAAGTGTCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTG 348
OY 241 tcaagtgtcaggtgcaggtcaggtcaggtgcaggtgcaggtgcaggtgcaggt 300
DB 349 GCAAGTGTCAAGTGTGCTCAAGTGTGCTCAAGTGTGCTCAAGTGTGCTCAAGT 408
OY 301 gcttcaggtgcaggtcaggtcaggtcaggtcaggtcaggtcaggtcaggtcaggt 328
DB 409 GGTCCAAGTGTGCTCAAGTGTGCTCAAGTGTGCTCAAGTGTGCTCAAGTGTGCT 468
OY 329 -----gtacaaatccttcaagaataatcaagtatcagatgtcaaatctaac 375
DB 469 AATTAATCTCATCTGTGCAAGCCCTCCAGCTGATGCAAGGATTCACATCTTAATTTAC 528
OY 376 gctgtatttaaacacagagtaacgaatactgttcaactatacaagaactcaaatcct 435
DB 529 GGTGATTATAAACACAGATGCAAAATTAATCTTGTTCATTAATAAACACCAATTAATCC 588
OY 436 caactcttgatttaactatacatatgttaacttggatgataataatcaatgattcaaa 495
DB 589 GAACCTTTGATTTAACCAATCATATGTTAACTTTGTGATTAATTAATGATGTTTCAAA 648
OY 496 tatttaattgataatgaagaataatlaaagaactataataataaactttatctt 555
DB 649 TATTATAATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 708
OY 556 gatttataatgaagcaaatlaaagaatgatagtgcataatgataatgataatccttc 615
DB 709 GATTATTATAAGCAAAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 768

OY 616 aatcttaaaatcgtgcaaaatgaattagacgtacttaaaaaacttgttcgagataaga 675
DB 769 AATCTTAATAATTGTGTCAAATGAATTAAGACGTCTTAAAAAACTTTGTTCGATTTAGA 828
OY 676 aaacattagacataatlaaagaataatgtagaagaataatggaagaattacataaaaaaat 735
DB 829 AAACATTACACAAATTTAAAGATTAATGTGGAAGAAATGGAAGATTATTAATAAAAAAT 888
OY 736 aaaaaaacatagaaaaataataatgaattaaatggaagaatgtagaagaataatgataa 795
DB 889 AAAACAACATTACCAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 948
OY 796 aataagaatcgaactaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 855
DB 949 AATAAGAAATGCAATTAATGAAGAAAGGAAAAAAATTTATCCAACTCATATGATCTT 1008
OY 856 tctattacaataaacaattagaagaagacacataatlaataagcgttttagaagaagcgt 915
DB 1009 TCTATTACAAATTAACAAATTAAGAAAGCACATTAATTAATTAAGCGTTTGAAGAAACGT 1068
OY 916 attgacacttlaaaaaaaatgaacaatlaaggaattactgtataagaataatgaatt 975
DB 1069 ATTGACACTTTAAAAAAATGAAGAAACATTAAGAAATTAATTAAGATTAATGAATTAAT 1128
OY 976 aaaaatccccaccgccaatttctggaataacacgaataatccttcttgaataagaacaa 1035
DB 1129 AAAAATCCCCACCAGCCCAATTTCTGGAATACCAAAATTCCTCTTGTATTAAGAACAAA 1188
OY 1036 aaaaatcaggaacacagaagaagaataaagaataatgccaataatlaaatttaatt 1095
DB 1189 AAAATCGAGAACACAGAAAGAAAATTAAGAAATTTGCCAAACTTTAAATTAATTAACATT 1248
OY 1096 gatagttattatctgtatccacttgaattagaatactattlaagagaagaataaataat 1155
DB 1249 GATAGTTTATTAATCTGATCCACTTGAATTAAGATTAATTAATTAATTAATTAATTAAT 1308
OY 1156 attgataatgagcaaggttga-----acaaagaatcaactgaaccatgaa 1206
DB 1309 GTTGATGTAACTTAATTAATCAAGATTCCTACGAAATTCGTTCAATTAATTAATTAATTA 1368
OY 1207 tatccaatggaattactatcccttgcataataagaataataaacttcaatcgttca----- 1260
DB 1369 TATCCAAATGATGTGATTAATCTTTACCACTCACGATTAATTAATTAATTAATTAATTAAT 1428
OY 1261 ----aatgaacttaattccttctgtgtatcaatcaatcaatcaatcaatcaatcaat 1317
DB 1429 GATTAATGATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1488
OY 1318 agtaaaacatatatacttgaatgaagaagaagaagaagaagaagaagaagaagaagaaga 1377
DB 1489 GAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1548
OY 1378 attaaatagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1437
DB 1549 ATGATGTTGAAGAAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1608
OY 1438 tcttaaatgataatacaaaagaataatgaagaataatcaatgaagaatcttagatagcaaa 1497
DB 1609 GATTATGAAGAAAGCAAAAAAGATTAATGAAGAAATTAATGAAGAAATTAATGAAGAA 1668
OY 1498 tccaataataatagatttaactaaatcgaagaagaagaagaagaagaagaagaagaaga 1557
DB 1669 TTTAATAATAATTTTGAAGAAAGTGTGATGAATAATAATTAATTAATTAATTAATTAAT 1728
OY 1558 aaagttagaagaacttacacacataataacttgcataatgaagaagaagaagaagaaga 1617
DB 1729 AATGTTGAAGAAACAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1788
OY 1618 ctggaagaattacaaagaagcctttaaataataggaagaattatcttlaagagatatagta 1677
DB 1789 GTTCAAAATTAATAAAGCTCTTTCATATCTTGAAGATTAATTAATTAATTAATTAATTAAT 1848

1678 gtgtgaaagaattlaaataatlaaataatlaaagaacaataagaataatgtgattgaa 1737
1849 TCAGAAAGATTTTATCATTTATTTATCTTGAAACCTGGCCCGAAGCTGATATATAA 1908
1738 acctgtgtgaaataatlaaataatgaagaacagcttlttgaaaaaataatctctaa 1797
1909 AAATTTAACAGAAATTAAGAGTAGAAGAAATTTCTAGAAAAAA--TTTAAA 1965
1798 gaagaataataacagaatgaataaatttgaagatctgacatgtgaaagatgaagtc 1857
1966 GAGCTTAACATTCAGCAATGGTCTTGAGAGTATCTATTTGTAATAATTAACAAGTA 2025
1858 caaaagtttatttaatlgaacaaatgacgaatlaaaaaagaatcgaattgttlaaa 1917
2026 CAAAAGTTTATTTAATTAATAAATAAGAGACTTAAAGAAAGATATATTTTAAAA 2085
1918 aatgtgaattaaacataatatacatgtcccaattcttaacaagaagaataagcaa 1977
2086 AATGCACACTAAAGTAGTATTCATGTACCAATTTATTAACCCAAAAATTAACCA 2145
1978 gaacctattatttaattgtgttgaataaagaataatgataaataagttcagct 2037
2146 GAACCATATTTATTTAATTTAATTAATAAAGAGATATTAATAATTAATTAATTA 2205
2038 aaggtgaatcatgtgataaagaataaataaataaataaataaataaataaataaata 2091
2206 AAAGTAAAGACATGTATTAAGAAAGAACAGCTGTCTTATCAAGTATTAACCAACCTTTA 2265
2092 tctgataatctcggaacatcaacacgaagaataaagaagaagaagaagaagaagaaga 2151
2266 GTTGCAGCAAGCAACCACTGAAGATGGGGTCACTCCACACACACATTTATCCCAATCA 2325
2152 ggaacaagaacagatctgtcttgaagaagatctgaagaagaagaagaagaagaagaaga 2204
2326 GGAAGAAACAGAAATGACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2385
2205 --agaacaataaagaacgaacaacgaatgacgaatgacgaatgacgaatgacgaatgac 2262
2386 ACCGTAAACATTAACATTTACCAACACACACATTCACCAACCAAGTAAGTAAGTAAGT 2445
2263 gtcccaacccacacgaacgaacgaataaataaataaataaataaataaataaataaata 2322
2446 GAAATTTCAATTAACATTAAGACTAATGACAAATTCACAAACCTTGACAAAGAAAGTAT 2505
2323 ctgaaataatcatgaattttaaatactataatgacgaataatctgtt 2382
2506 CTTAAGAAATTAATGATGATTTTAACTAATCATATATATGCTCAATAATATTTTAACTA 2565
2383 tcaactcaactaagaagaagaataatlaaaacaataaataatlaaaagaagaagaaga 2442
2566 TCAACTCTGTATGACCAAAAAATTTATGAGTATATATATCTTCCAGAAAGAA 2625
2443 agaaataatgaatcatgtgacatctgttataataaataaataaataaataaataaata 2502
2626 AATGAATTTAAATCAATGTCATTAAGATTTTATTTATTTTAAATTAACATTAATCT 2685
2503 gtaattattctatgttgaatgacgaataaataaataaataaataaataaataaataaata 2562
2686 GCATGTATTTCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2745
2563 tatgaagaagaataatgttgaatattataaataaataaataaataaataaataaataaata 2622
2746 TATCAAAAGAAATGATTTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 2805
2623 ttaattgaagaagaagaataaataaataaataaataaataaataaataaataaataaata 2676
2806 TTTTGTAGAGAGCAAAACAAATTAATGATGATGATGATGATGATGATGATGATGATGAT 2865
2677 tcaatgaacacatcatcatcatcatcatcatcatcatcatcatcatcatcatcatcatcat 2717
2866 ACCGTAATTAATGCTCTCAATCCGACCTCAAGTAAATTTCCCAAAACCAACATCAATCA 2925
2718 -----agtaagtga 2727

2926 TCCTCTACCAATACCCAAATGCTGTAGCTGTATCATCTGCTGCTGCTGCTGCTGCTGCTG 2985
2728 aatgaatgatacatcatcatcatcatcatcatcatcatcatcatcatcatcatcatcatcat 2787
2986 AGTCATGATCCCTTAACAGTATTTGTCTATTAATGATGATGATGATGATGATGATGATGAT 3045
2788 ttagctcttgaaaaaacaataaataaataaataaataaataaataaataaataaataaata 2844
3046 TTAATCTTGGAAATTAACCTTAAGTACCTTAATCCATTAACCTTTCTACACAGAGATG 3105
2845 gaaataatlaagaagaatlaagaagaatlaagaagaatlaagaagaatlaagaagaatlaaga 2904
3106 GAAATAATTTATGAGATATTTTAAATAATGATGATGATGATGATGATGATGATGATGATG 3165
2905 aatttgaataatctgaagctgtatcatcatcatcatcatcatcatcatcatcatcatcatcat 2964
3166 CAATTCGTAAATCTAATTTCAAAAGTAAATTAAGGTTGA--CCGAAACCAAAAAAT 3222
2965 aatltgaagaagaatlaaataaataaataaataaataaataaataaataaataaataaata 3024
3223 GCATTAATGATGAATTTAAATAATTAAGATACCTTTACAGTTATCATTTGATTTATAT 3282
3025 aataaataaataaataaataaataaataaataaataaataaataaataaataaataaataa 3084
3283 AATTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3342
3085 aaaaatgaataaataaataaataaataaataaataaataaataaataaataaataaataa 3144
3343 AAATGCAAAATTAATAAACTTACTTTTAAAGAACAAATTAATTAATTAATTAATTAATTA 3402
3145 ctlaataacccaagaacgtatcatcaataaataaataaataaataaataaataaataaataa 3204
3403 CTTAATTAACCATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3462
3205 gctgaatagcagaac 3264
3463 GCTGAATATACCAAACTGAAACACATTTGAAAGAACCAAAATTAATTTGAAAGAACTTAT 3322
3265 aaagagctgttgaataatlaaataaataaataaataaataaataaataaataaataaataa 3324
3523 AAAGACTTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3382
3325 tcaattcaacagaagaataatlaaataaataaataaataaataaataaataaataaataaata 3384
3583 TCAATTAACCAAGATATTTATGCAATTTAGAAAAATTTAGAGATTAAGTAAATA 3642
3385 gaaagaataatlaagaagaatlaaataaataaataaataaataaataaataaataaataaata 3444
3643 GATGAAGAAACCTCAATGATTAATTTAATTTAGAAAGCAAAATTAATCTTTATCAAGT 3702
3445 ggaattacatcatcatcatcatcatcatcatcatcatcatcatcatcatcatcatcatcat 3504
3703 GGATTTACATATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3762
3505 aatttccaaatgaataaataaataaataaataaataaataaataaataaataaataaataa 3564
3763 AATTTCCAAGTAAATTAATTAAGAAAGTTAACCAAGCTTTAAATTTTTCGAAAAATTTT 3822
3565 ctccagaagaagaagaat-----gt 3585
3823 CTCGCCAAGAACAAAGTTTACACAGTTGTAATTCACCTCAACCTCAACCTCAACCTCAACCT 3882
3586 gaaacagttgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 3645
3883 CCATTTATCTGTAGAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 3942
3646 gcatcaactcatgtgaagaagaatgaagaagaatgaagaagaatgaagaagaatgaagaaga 3705
3943 ACTTCAGGCTCTTTATTAATTAAGATTAACCAAGTATGATTAATTAACCAATTAATTAAT 4002
3706 gaagttagatgaatlaaataaataaataaataaataaataaataaataaataaataaataaata 3765

[illegible]

Dh 2038 CATGTACCAATATTTATTAACCAAAATTAACAGAACCATATATTATTATTTATTTA 2097
Qy 2002 aaaaaagaattgataaatttaaagaatgcttaagcttaagatgataatgataa 2061
Db 2098 AAAAAAGAAAGTATTAATTAATTAAGAAATTTATTAACAAAAGTAAACATATGTTAAAGAAA 2157
Qy 2062 gaaaaaanaaacaataaagaagctgaatcgtgaattcgtgaacccaacgaagga 2121
Db 2158 GAACAAGAGTGTCTTATCAAGTATTACACAACCTTTAGTTGACAGACGCAAC--AACT 2214
Qy 2122 gaataaagaagaagaacatacaaaacctggaacaagaagatcgtcttagaaga 2181
Db 2215 GAGATGGGGCTCCTCCACACACATTTATCCCAATTCAGAGAAACAGAGTAACAGAA 2274
Qy 2182 gattagtgacagacagaacacagaacaaaaaagaacacaccagtlaccagtaaca 2241
Db 2275 GAACAGGAAGAAACAGAAACAGTAGACACACAAACGGTAACTAATTAACATTAACA 2334
Qy 2242 gtaacagaagaacaaagacaaagtcaccaacacacagcaacagtaataaactgaa 2301
Db 2335 CCAAAAGAGTAAAGTGTGAAATTTCAATTAAGACATTAAGTAAATGCAATTCACAA 2394
Qy 2302 aatgcttccaatagatatacttgaaaaaatalaagaatttaaatcttaataata 2361
Db 2395 GCGTTGACAAAAACAGTTATCTAAGAAATTAAGATGAATTTTAACTAATCATATATA 2454
Qy 2362 tctgaataatatttggcttccacactgaacataagaagaagaatlataaacatat 2421
Db 2455 TCGCATTAATTAATTTTATGATCAAACTAGATGACCAAAATTAATTAAGGATATAT 2514
Qy 2422 aaaaatacaagaagaagaagaacaaatlaaagtlcagtcacatagactatgctt 2481
Db 2515 AATCTTACCCAGAGAAAGAAATGAATTAATCAATCATGATGATCATTAATTTATTT 2574
Qy 2482 aatacaaaaatacaatccgtaactgtaactgtaactgtagatgtaaaacatgta 2541
Db 2575 AATATTCAAAATTAACATCTGATGATGATGATGATGATGATGATGATGATGATGAT 2634
Qy 2542 tcaacactatattagaataattatagaagaagaagaatgcttgaattatataaactaa 2601
Db 2635 CAACATCTCTTTTGTGATTAATTAATCAAAAGAAATGATTTATTTATTAATTAACATAA 2694
Qy 2602 gataatgacaaaattaaattatattagaagaagcgaaaaaa-----gtaccacact 2655
Db 2695 GAGGAAATACATCAATAAATTTATGAGAGCAAAACAAATTAATCGAATCATCATCT 2754
Qy 2656 gtaaaaaaccttcaagctcaatgaatgaacacatcaatcaacactcaagaataacc 2715
Db 2755 ACATCCAGTCTGGAATTAACACCGTAATTAATTAATTAATTAATTAATTAATTAAT 2814
Qy 2716 ga----- 2717
Db 2815 CAAMACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2874
Qy 2718 -----agtaagtgacaaatgtaatacatcaatccttaacaaattgataatagt 2766
Db 2875 GGTCTGCTGTGAGTGAAGAAGTATGATCCCTTAACAGTATTTGCTTAATTAATTAAT 2934
Qy 2767 ttaaaattcttgaaacataatgagcttggaagaacacaaataatatacagaagata 2826
Db 2935 TTGAAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2994
Qy 2827 ---ataagtgacaaagaatgtaagaactttatagaagaatataaagaatgtaaga 2883
Db 2995 ACCATTCTTACACAGAGTGAAGAAATTTTATGAGAAATTTTAAATAATTAATTAATTAAT 3054
Qy 2884 ttcttaaatgaatcttcttaacaaatttgtaaaatccttaagaagctgaatgataatcaatc 2943
Db 3055 TATTTTAATGATGATTAACAACATTCGTAATAATTAATTAATTAATTAATTAATTAATTAAT 3114
Qy 2944 aatgatgaatcaaaagaagaattagaagaatataatataaataaataaactta 3003
Db 3115 A---CCGAACACACAAAAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3171

Qy 3004 cagttatcatttgattatataataataaataaattgaagaatatttgataaa 3063
Db 3172 CAGTTATCATTTGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3231
Qy 3064 aagaagaagcttgtaataataaataaataaataaataaataaataaataaataaataaataa 3123
Db 3232 AAAAAAGACTTGGCAGACAAATGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3291
Qy 3124 ttgaatcaaaatttgaaatcacttaataaataaataaataaataaataaataaataaataaataa 3183
Db 3292 TTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3351
Qy 3184 ttcttcaaaaaaagaagaatgtaagaagaagaagaagaagaagaagaagaagaagaagaaga 3243
Db 3352 TTCTTTAACAATAAAGAGCTGAATTAAGCAAGAACTGAATAACATTAAGAAACACA 3411
Qy 3244 aaaaataatlgaaacattataaagaagcttgtaaaatataatgtaagtaacatcca 3303
Db 3412 AAAAAATTTATGAAACATTAATAAGACCTTGTAATTAATTAATTAATTAATTAATTAATTAAT 3471
Qy 3304 ttaaaaacttgaatgaaatcaatccaagaagaataatgacagttagaagaac 3363
Db 3472 TTAAAACTTTAAGTGAAGTATCAATTAACAGAGATTAATTAATTAATTAATTAATTAATTAAT 3531
Qy 3364 tttaagttatgaatgaatgaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 3423
Db 3532 TTATGAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3591
Qy 3424 aaaaataatlgaaacattataaagaagcttgtaaaatataatgtaagtaacatcca 3483
Db 3592 AATTTATCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3651
Qy 3484 aaaaataatlgaaacattataaagaagcttgtaaaatataatgtaagtaacatcca 3543
Db 3652 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3711
Qy 3544 ttgaatcttcaaaaaatttccccaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 3582
Db 3712 TTAAAACTTTAAGTGAAGTATCAATTAACAGAGATTAATTAATTAATTAATTAATTAATTAAT 3771
Qy 3583 -----gtgcaacagcttgtaagtgaaatgtaagtgacacacatlagaa 3624
Db 3772 CAACGATGTAACCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTCATCT 3831
Qy 3625 caaagtcaacaaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 3684
Db 3832 AAAGAGAAACACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3891
Qy 3685 acatcaaaaatgctgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 3744
Db 3892 CAATTTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3951
Qy 3745 tccgaagaagattatgataatgaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 3804
Db 3952 TCCGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4008
Qy 3805 gtaattgataaacttcttaaaattgaagaagaagaagaagaagaagaagaagaagaagaagaaga 3864
Db 4009 ACATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4068
Qy 3865 ttgaagtgatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 3924
Db 4069 TTAGCTGAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4128
Qy 3925 gtaattgataaacttcttaaaattgaagaagaagaagaagaagaagaagaagaagaagaagaaga 3984
Db 4129 TTAAATTTGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4188
Qy 3985 ttgaatcgaatgaatccatataaagaatgaagaagaagaagaagaagaagaagaagaagaagaaga 4044
Db 4189 TTGAATCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4248

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2001, 15:51:20 ; Search time 53.94 seconds
(without alignments)
4020.169 Million cell updates/sec

Title: US-09-269-874-3
Perfect score: 8424
Sequence: 1 MKIIFLCSEFLFIINTOCV.....SNFLGISFLILMLILYSFI 1639

Scoring table: BLOSUM62 , Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP:REMBL_16:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.rodent:*
13: sp.unclassified:*
14: sp.vertebrate:*
15: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5283.5	62.7	1704	5	Q9TZT4	Q9TZT4 plasmodium
2	5162.5	61.3	1694	5	Q9TZT5	Q9TZT5 plasmodium
3	5162.5	61.3	1694	5	Q9NHX1	Q9NHX1 plasmodium
4	5107.5	60.6	1720	5	Q25922	Q25922 plasmodium
5	3093.5	36.7	1751	5	Q26194	Q26194 plasmodium
6	3003.5	35.7	1087	5	Q25961	Q25961 plasmodium
7	2951	35.0	1726	5	Q02569	Q02569 plasmodium
8	2801	33.3	539	5	Q9TYG1	Q9TYG1 plasmodium
9	2798	33.2	539	5	Q25972	Q25972 plasmodium
10	2796	33.2	539	5	Q25966	Q25966 plasmodium
11	2795	33.2	539	5	Q25973	Q25973 plasmodium
12	2794	33.1	539	5	Q25976	Q25976 plasmodium
13	2791	33.1	539	5	Q25981	Q25981 plasmodium
14	2789	33.1	539	5	Q25984	Q25984 plasmodium
15	2788	33.1	539	5	Q25971	Q25971 plasmodium
16	2532	30.1	1787	5	Q25645	Q25645 plasmodium
17	2530	30.0	1791	5	Q9U6D4	Q9U6D4 plasmodium
18	2509	29.8	1785	5	Q25685	Q25685 plasmodium
19	2494.5	29.6	1766	5	Q25668	Q25668 plasmodium

20	2146	25.5	651	5	Q25924	Q25924 plasmodium
21	2119.5	25.2	652	5	Q25923	Q25923 plasmodium
22	2101	24.9	400	5	Q03999	Q03999 plasmodium
23	1818	21.6	360	5	Q9NAT3	Q9NAT3 plasmodium
24	1778	21.1	569	5	Q25975	Q25975 plasmodium
25	1772	21.0	569	5	Q25969	Q25969 plasmodium
26	1772	21.0	569	5	Q25977	Q25977 plasmodium
27	1768	21.0	569	5	Q25974	Q25974 plasmodium
28	1768	21.0	569	5	Q25979	Q25979 plasmodium
29	1767	21.0	569	5	Q25983	Q25983 plasmodium
30	1765	21.0	569	5	Q25967	Q25967 plasmodium
31	1758.5	20.9	570	5	Q25968	Q25968 plasmodium
32	1753	20.8	569	5	Q25982	Q25982 plasmodium
33	1749	20.8	569	5	Q25970	Q25970 plasmodium
34	1748.5	20.8	570	5	Q25972	Q25972 plasmodium
35	1748	20.8	569	5	Q25978	Q25978 plasmodium
36	1745	20.7	569	5	Q25980	Q25980 plasmodium
37	1696	20.1	336	5	Q25865	Q25865 plasmodium
38	1670	19.8	356	5	Q9TZU5	Q9TZU5 plasmodium
39	1649	19.6	350	5	Q9TZU2	Q9TZU2 plasmodium
40	1636.5	19.4	357	5	Q9TZU7	Q9TZU7 plasmodium
41	1634.5	19.4	357	5	Q9TZU8	Q9TZU8 plasmodium
42	1625.5	19.3	360	5	Q9TZV2	Q9TZV2 plasmodium
43	1625.5	19.3	363	5	Q9TZU9	Q9TZU9 plasmodium
44	1622	19.3	344	5	Q9TZU1	Q9TZU1 plasmodium
45	1616.5	19.2	362	5	Q9TZV3	Q9TZV3 plasmodium

ALIGNMENTS

RESULT	ID	Q9TZT4	PRELIMINARY:	PRT:	1704 AA.
AC	Q9TZT4	01-MAY-2000 (TREMblrel. 13, created)			
DT	01-MAY-2000 (TREMblrel. 13, last sequence update)				
DT	01-JUN-2000 (TREMblrel. 14, last annotation update)				
DE	MEROZOITE SURFACE PROTEIN 1 PRECURSOR.				
GN	MSPI.				
OS	Plasmodium falciparum.				
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
OX	NCBI_TaxID=5833;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	STRAIN-HN2;				
RC	Jiang G., Liu R.Z., Daubenberger C.A., Pluschke G.;				
RL	Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.				
DR	EMBL: AF062349; AAC72885.1; -				
DR	InterPro: IPR000561; -				
DR	Pfam: PF00008; EGF; 1.				
KW	Merzoite.				
SO	SEQUENCE	1704 AA;	193762 MW;	385526DDA56FD1D CRC64;	
Query Match		62.7%;	Score 5283.5;	DB 5;	Length 1704;
Best Local Similarity		62.1%;	Pred. No. 2.1e-187;		
Matches 1075;		Conservative 219;	Mismatches 317;	Indels 121;	Gaps 26;
OY	1	MKIIFLCSEFLFIINTOCVTHESVOELVKLEALDELVDLVLTGYSIFQEKWLVNGTSGT	60		
DB	1	MKIIFLCSEFLFIINTOCVTHESVOELVKLEALDELVDLVLTGYSIFQEKWLVNGTSGT	60		
OY	61	AVTSTPGSK--GSVASGSGSVASGSGSVASGSGSVASGSGSRRTPSDNS	117		
DB	61	AVTSTPGSKSGSGSVASGSGSVASGSGSVASGSGSVASGSGSRRTPSDNS	117		
OY	118	SDSAKSYADIKHRYRNLITIKELYPQLPDLTNHMTLTDNTHGFRYLLDGEETNEL	177		
DB	118	SDSAKSYADIKHRYRNLITIKELYPQLPDLTNHMTLTDNTHGFRYLLDGEETNEL	177		
OY	178	LYKLAFYFDLLRAKLVDCANDYCOIPENLKRANEDLVKLKLVFGYRKPDPNKKDNNGK	237		

```

Db 178 LYKLFYDYLRAKINDACANSYCOIPFNLIKIRANELDVLAKKIVGRRKPLDNIKDNVKG
Qy 238 MEDYIKKKKTITENINELIESSKTIIDKNKATKEEKKKIYQAOYDSTYNKOLEEPAHN 297
Db 238 MEDYIKKKKTITANINELIESSKTIIDKNKADNEGGKKLIYQAOYDSTYNKOLEEPAHN 297
Qy 298 LISVLEKRIIDTLKKNENIKELLDKINEIKNPPANSNGTPTLIDKNKIEEHEKEIKEI 357
Db 298 LISVLEKRIIDTLKKNENIKELLDKINEIKNPPANSNGTPTLIDKNKIEEHEKEIKEI 357
Qy 358 AKTIKFNIDSFTDPLELEYLREKKNIDISAKVE--TKESTEN-ETPGVITYPLSYN 414
Db 358 AKTIKFNIDSFTDPLELEYLREKKNIDISAKVE--TKESTEN-ETPGVITYPLSYN 417
Qy 415 DINNAL--NELNSFGDILNPPDYTKEPSKNITPDN-ERKKFINIEKIKIEKKI--- 467
Db 415 DINNAL--NELNSFGDILNPPDYTKEPSKNITPDN-ERKKFINIEKIKIEKKI--- 476
Qy 468 -ESDKSYEDRSKSLNDITKEYEKLNEIYDSKFNNNIDLTNEFKMGKRYSYVEKLTN 526
Db 477 KEONKKLIEDYERS---KDYEELEKFEYEMKFNNNFDKDVODKIFASRYTYNEKQRY 532
Qy 527 HNTFASYSKHNLEKTLKALKYMEDYSLRNIVAKELKYTKNLISKIENETIYENIK 586
Db 533 NKSSSSNNYSYVQOKLKALSYLEDYSLRKGISSEKDNHYATLTGTEADIKLIEBIEIK 592
Qy 587 KDEOLFEEK---ITKDNKPKDEKILEVSDIYKVOYOKVLLMNKIDELKTKQILKJNEL 643
Db 593 SSEKILKKNKGLTHSNANAS---LEVSDIYKVOYOKVLLMNKIDELKTKQILKJNEL 648
Qy 644 KHNHVPNSYKQENKQEPYLLIVLKEKIDKLVPMKVESLINERKNITKEGOSDNSEP 703
Db 649 KDSIHVPNITYKQNKPEPYLLIVLKEKIDKLVPMKVESLINERKNITKEGOSDNSEP 699
Qy 704 STEGITOQATTKPQOQGSALIEGDSYQAOQOQAOQ-----PPVYPPVPEAAQYPT 757
Db 700 ITQPLVASETTEDEGGHSTHLSQSGETEVEETEETEVEGHTTIVTITLPTPQ---PS 756
Qy 758 PPAPV-----NKTENVSKLDYLEKLEFPLNTSYCHKILVSHSMNKKILKQ 806
Db 757 PPKEVYVVENSTIEHKSNNQALTYVYLKIDLEFLTKSYCHKILVSHSMNKKILKQ 816
Qy 807 YKITEESKLSQCDPLDLFNIONNIPVMSMPDLSNLSQLEMEIYEMVONLYKL 866
Db 817 YNLPEEKNEKSCDPLDLFNIONNIPVMSMPDLSNLSQLEMEIYEMVONLYKL 876
Qy 867 KONDKIKULLEBAKV-----STSVKTLSSS-----MPLSTLPDQKPEVS 908
Db 877 KEENHIKILKEOKOITGTSSTSPGNTVYMAQSATHSNQOOSNASTWTONGAVAS 936
Qy 909 AND---DTSHS---TNLNSLKLEFNILSGKNKNITQOL-IGQKSENEFEKILKSD 960
Db 937 SGPAVESHPPLVYLSNDLKGIVSLNKGKTKVNPPLITSTTEKEFELENKND 996
Qy 961 TFYNESTFNPYKSKADILNSINDESKRRLLEEDINKLKTQLSFDLYNKKYKLEERFD 1020
Db 997 TYFNDNDIQFVKSNSKVITGLT-ETQKNAIENDEIKKLDIQLSDLYNKKYKLEERFD 1055
Qy 1021 KKKVYGVKTKMOIKKLTLLKEQLESKLSLNPKHVLQNFVFPENKKKAELAEIENLEEN 1080
Db 1056 KKKVYGVKTKMOIKKLTLLKEQLESKLSLNPKHVLQNFVFPENKKKAELAEIENLEEN 1115
Qy 1081 TKILLKHKGLVKKYNGSSPLKTLSEESIGTEDVYASLENFKVLSKLEKIKOVLNLEK 1140
Db 1116 TKILLKHKGLVKKYNGSSPLKTLSEESIGTEDVYASLENFKVLSKLEKIKOVLNLEK 1175
Qy 1141 KKLSTLSSGLHLIAELAEVLKKNKNTGNSPENNTPDNNALLESYKFLPREGDVAITYS 1200
Db 1176 KKLSTLSSGLHLIAELAEVLKKNKNTGNSPENNTPDNNALLESYKFLPREGDVAITYS 1234
Qy 1201 E-----SGSDTLEQSQPKRPASTHVGASNTTTSNDVDEVDVITVPIF 1246
Db 1235 PPOPDVTPSPLSVRSVSGSGSTKETQIPTSGSLTTELOOVVQSONYDEEDDSLVLPITF 1294

```

```

Qy 1247 GSEEDYDILGOVYTGAVTPEVSDINILSKITENEYEVLYKPLAGVRSKLOENNYMT 1306
Db 1295 GSEEDYDILGOVYTGAVTPEVSDINILSKITENEYEVLYKPLAGVRSKLOENNYMT 1353
Qy 1307 FNVNKKDILNSRFNKNRENFKNVLESDLIPYKDLTSSNVKVPKFKELNKRDRFLSSYN 1366
Db 1354 FNLNLDILNSRLKRRKFFLDVLESDLMQFKHISSENEYITDESFKLNSDOKNTLLSKY 1413
Qy 1367 YIKOSIDIDINPANDVGLGYKILSKYSKSDLSIKKYI-----NDK 1407
Db 1414 YIKESVENDIKFPAQGISYERKYLAKYKDDLESIKKVIKEKEPPSPPTPPSPAKTDE 1473
Qy 1408 OGENSEKYLPFLNNTETLKTYVNDKIDLEFVILHEAKVLWYTEKSNVEKIKELNYLKTIO 1467
Db 1474 OKESKELPFLNNTETLKTYVNDKIDLEFVILHEAKVLWYTEKSNVEKIKELNYLKTIO 1533
Qy 1468 DKLADFKKNNFVGIADISTDYNNHNNLLTKFLSTGCVFENLAKTYVLSMLDGNLQMLNI 1527
Db 1534 DKIDLFKMTNDFEALIKLINDYTKKMDLGLSTGLV-ONPPNTITSKILGKFDQMLNI 1592
Qy 1528 SOHCYVKKQCPQNSGCFPHHDERECKLLNWKQBGKCVENPNTCENNNGCDADAKC 1587
Db 1593 SOHCYVKKQCPQNSGCFPHHDERECKLLNWKQBGKCVENPNTCENNNGCDADAKC 1652
Qy 1588 TEEDSGSNKKITCECTKPDSPYPLFDGIFCSCSSNFGISFLILMLILYSFI 1639
Db 1653 TEEDSGSNKKITCECTKPDSPYPLFDGIFCSCSSNFGISFLILMLILYSFI 1704

RESULT 2
Q9TZT5 PRELIMINARY; PRT: 1694 AA.
AC Q9TZT5.
ID Q9TZT5.
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE MEMOZOITE SURFACE PROTEIN 1 PRECURSOR.
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HN1;
RA Jiang G., Liu R.Z., Daubenberg C.A., Pluschke G.;
RL Submitted (Apr-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF062348; AAC72884.1; -.
DR InterPro; IPR000561; -.
DR Pfam; PF00008; EGF_1.
KW Merozoite.
SQ
SEQUENCE 1694 AA: 192794 MW: 84CFCD0E709F5673B CRC64;

Query Match 61.3%; Score 5162.5; DB 5; Length 1694;
Best Local Similarity 61.1%; Pred. No. 6,1e-183;
Matches 1054; Conservative 226; Mismatches 328; Indels 117; Gaps 27;

```

QY 241 YKKNKTTIENINLEIESKKTIDKNNKATEEKKKLYQAQYDLSYKNOLEBAHNLIS 300
 D 235 YKKNKTTIANININLEIESKKTIDKNNKADNEEKKKLYQAQYDLSYKNOLEBAHNLIS 294
 QY 301 VLEKRIOTLKKNEIKELLOKINEIK-NPPFANGNTPNLTLD-KNKKIEHEKEIKIA 358
 D 295 VLEKRIOTLKKNEIKELLOKINEIK-NPPFANGNTPNLTLD-KNKKIEHEKEIKIA 354
 QY 359 KTIKFNIDSLFTDPLELEYLREKNKIDISAKVE--TKESTERN-EXPNGVTPYLSYND 415
 D 355 KTIKFNIDSLFTDPLELEYLREKNKIDISAKVE--TKESTERN-EXPNGVTPYLSYND 414
 QY 416 INNAL--NELNSGDLINPDYTKEPSKNIYTON-EAKKINIEKIKIEKKI----- 467
 D 415 INNAL--NELNSGDLINPDYTKEPSKNIYTON-EAKKINIEKIKIEKKI----- 467
 QY 468 ESDKSYEDSKSLNDITKEVEKLINLEYDSEKFNNDIDLTNFEKMMGKRSYKVEKLTJH 527
 D 474 ESDKSYEDSKSLNDITKEVEKLINLEYDSEKFNNDIDLTNFEKMMGKRSYKVEKLTJH 529
 QY 528 NTFASYSKHNLEKLTALYKMEYSLRNIVYERKELKYKNIISKIENEJETEVENIK 587
 D 530 NTFASYSKHNLEKLTALYKMEYSLRNIVYERKELKYKNIISKIENEJETEVENIK 589
 QY 588 DEBOLFEEK--ITKDEKRPDEKILEVSDIYKVOYKVLNKKIDELKKTOLILKNELEK 644
 D 590 DEBOLFEEK--ITKDEKRPDEKILEVSDIYKVOYKVLNKKIDELKKTOLILKNELEK 645
 QY 645 HNIVPVSYSKOEKOEPPYLLVLEKKEIDKLVPMKPVESLINEKKNKITEGQSDNSEPS 704
 D 646 HNIVPVSYSKOEKOEPPYLLVLEKKEIDKLVPMKPVESLINEKKNKITEGQSDNSEPS 704
 QY 705 TEGITQOATTKPGQQAAGSALSGSVOAQAQOEKQADPVPVPEAKAQPVPAPV-- 762
 D 697 TEGITQOATTKPGQQAAGSALSGSVOAQAQOEKQADPVPVPEAKAQPVPAPV-- 762
 QY 763 TEGITQOATTKPGQQAAGSALSGSVOAQAQOEKQADPVPVPEAKAQPVPAPV-- 762
 D 754 TEGITQOATTKPGQQAAGSALSGSVOAQAQOEKQADPVPVPEAKAQPVPAPV-- 762
 QY 814 ESKLSQDPLDLFNIQNNIPVMSMPDSLNLSQLEMEIEREKEMCNLYKLDNDKIK 873
 D 814 ESKLSQDPLDLFNIQNNIPVMSMPDSLNLSQLEMEIEREKEMCNLYKLDNDKIK 873
 QY 874 NLEBAKV-----STSVKTLSSSS-----MPLSLTPDOKPEVSAAND---D 912
 D 874 NLEBAKV-----STSVKTLSSSS-----MPLSLTPDOKPEVSAAND---D 912
 QY 913 TSHS-----TNANSLKLFENLISLGGKNNIYOEL-IGQKSESENYEKLKOSDPFYNSEF 967
 D 913 TSHS-----TNANSLKLFENLISLGGKNNIYOEL-IGQKSESENYEKLKOSDPFYNSEF 967
 QY 968 TNPVSKADADINSINDESKRRLLEEDINKLKTLOLSFDLYNKKYKLEFEDKKKTYGK 1027
 D 994 TNPVSKADADINSINDESKRRLLEEDINKLKTLOLSFDLYNKKYKLEFEDKKKTYGK 1027
 QY 1028 YKMOIKKLTLEKOLESKLNSLNPKHVLONFVFPENKKKAEALAEETENTLENTKILKH 1087
 D 1053 YKMOIKKLTLEKOLESKLNSLNPKHVLONFVFPENKKKAEALAEETENTLENTKILKH 1112
 QY 1088 YKGLVKKYNGSSPLKLTSESIOTEDNYASLENFYKLSKGLKXDLNLEKKKSLYS 1147
 D 1113 YKGLVKKYNGSSPLKLTSESIOTEDNYASLENFYKLSKGLKXDLNLEKKKSLYS 1147
 QY 1148 SGLHLLIAELKEVINKNNKYTGNSENNTPDYNNALESYKRLPEGTDAATJVSSE----- 1201
 D 1173 SGLHLLIAELKEVINKNNKYTGNSENNTPDYNNALESYKRLPEGTDAATJVSSE----- 1201
 QY 1202 -----SGSDTLQEQOPKRPASTHVGAESNTJTSQNVDEVDVAVIIPFGSEEDY 1253
 D 1232 PSLPVSIVRSGSGSGKKEBTQIPTSGSLTLELQOVVOSONVDEEDSLVVLPIFGSEEDND 1291

QY 1254 DDLGQVVTGEAVTSPVIDNINILSKITENEVEVLYLKLGVYRSLLKKOLENNVFNENVKD 1313
 D 1292 EYLDQVVTGEAVTSPVIDNINILSKITENEVEVLYLKLGVYRSLLKKOLENNVFNENVKD 1350
 QY 1314 ILNSFRNKRNFNKLLEEDLPRYKDLTSSNVVADKPFNLKERKDFLSSYNYIKSD 1373
 D 1351 ILNSFRNKRNFNKLLEEDLPRYKDLTSSNVVADKPFNLKERKDFLSSYNYIKSD 1410
 QY 1374 TDINFANDVLGYKILSEKYSIDSLIKKY-----NDKOGENEKY 1414
 D 1411 TDINFANDVLGYKILSEKYSIDSLIKKY-----NDKOGENEKY 1414
 QY 1415 LPLNTIETLYKYNKIDLEFVHLEKVLNVTYKESNVEKIKELNVTYKIDOKLDFK 1474
 D 1471 LPLNTIETLYKYNKIDLEFVHLEKVLNVTYKESNVEKIKELNVTYKIDOKLDFK 1530
 QY 1475 KNNFVGIADLSTDYNNHNLTKFLSTGMVFNELAKTVLSMLDGNLQGMNISQHOVC 1534
 D 1531 KNNFVGIADLSTDYNNHNLTKFLSTGMVFNELAKTVLSMLDGNLQGMNISQHOVC 1589
 QY 1535 KOPNSGCFRHLDERECCLNLYKOGDKCVENPNTENNNGCDADAKTEEDSGS 1594
 D 1590 KOPNSGCFRHLDERECCLNLYKOGDKCVENPNTENNNGCDADAKTEEDSGS 1649
 QY 1595 NGKRTCECTKPDSPPLFDGIFGSSNPLGTSFLILMLILYSPL 1639
 D 1650 NGKRTCECTKPDSPPLFDGIFGSSNPLGTSFLILMLILYSPL 1694
 RESULT 3
 Q9NHX1 PRELIMINARY; PRT: 1694 AA.
 AC Q9NHX1;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE MAJOR MEROZOITE SURFACE ANTIGEN.
 GN GP195.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_taxid=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FCCL/HN;
 RA Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M.;
 RT "Molecular cloning and sequence analysis of major merozoite surface
 RT antigen(gp195)gene of Plasmodium falciparum isolate FCCL/HN."
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF218248; AAF27526.1;
 DR InterPro: IPR000561;
 DR InterPro: IPR003247;
 DR Pfam: PF00008; EGF 1;
 DR ProDom: PD001527; -; 1.
 KW Merozoite.
 SQ SEQUENCE 1694 AA; 192766 MW; B51634A49E0F6728 CRC64;
 Query Match 61.3%; Score 5162.5; DB 5; Length 1694;
 Best local similarity 61.1%; Pred. No. 6.1e-183;
 Matches 1054; Conservative 226; Mismatches 328; Indels 117; Gaps 27;

181 LNFYDLRAKLVNDVANDYCOIPENLIRANLDELVLKLVGRKPLDNKIDNVKMD 240
175 LNFYDLRAKLVNDVANDYCOIPENLIRANLDELVLKLVGRKPLDNKIDNVKMD 234
241 YIKNNKTIENINLIEESKKTIDKNNKAREEKKKLYQAOYOLSTYNNOLEAHNLIS 300
235 YIKNNKTIENINLIEESKKTIDKNNKAREEKKKLYQAOYOLSTYNNOLEAHNLIS 294
301 VLEKRIOTLAKNENIKELLDKINEIK-NPPRANGNTPNLTLD-KNKKIEHEKEIEIA 358
295 VLEKRIOTLAKNENIKELLDKINEIK-NPPRANGNTPNLTLD-KNKKIEHEKEIEIA 354
359 KTIKENIDSLFTDPLEEYVIREKNKIDISAKVE--TRKSTERN-EYPNGVTPYLSYND 415
355 KTIKENIDSLFTDPLEEYVIREKNKIDISAKVE--TRKSTERN-EYPNGVTPYLSYND 414
416 INNAL---NELNSFGDLPNPDYTKERSKNITPYN-ERKFINIKERKIEKKKT---- 467
415 INNSLADNDKNSYGDLMNP-DYKELINEKITDNEKRIEINNIKQIDLEENINHTK 473
468 ESDKSYEDRSKSLNDITKEYEKELINEYDSEKFNNDITLTFEEMMGKRSYKYEKLTTH 527
474 ESKKLEEDLEKES---KROYELLEKEFEKFPNNNDKVDYDKIFESARTYNEKOKYN 529
528 NTFASYENSKHNLEKTLKALYMEDYSLRNIYVEKELKYKNLISKEIETLVENIKK 587
530 NKFSNNNSYVNOYKTLKALYSLRKGISEKDFNHYTTLTGLEADIKKLTLEIKS 589
588 DEBOLPEK---ITKDNKDEKLEYSOYKVOYVOLLNKLDELKTOILIKNVELK 644
590 SENKILEKNEKGLSHSNAS---LEVSDIVKLOVOKVLLIKLEDEKIELFLKNOQK 645
645 HNTHVPSYKOEKNKOEYVILVILKLEKIDKLVMPKVESILNEEKNKITEGSDNSEPS 704
646 DSHVPIYKPKOKPEYVILVILKLEKIDKLVMPKVESILNEEKNKITEGSDNSEPS 696
705 TEBELTGATTPGQOAGSALSDSDYQAOEOKOAPPVPPVPEAKAOPVPPAPV-- 762
697 TOPLVASEETTEGSGHSTHLSOGETEETEEVGHHTTATITLPPG---PSPPKEVK 753
763 -----NNKTEVNSKLDYLEKLYEFLNTSYCHKYYIIVSHSTNMEKILKOKITKEE 813
754 VENSIEHKSNDNSOALTKYVLLKLEKIDKLVMPKVESILNEEKNKITEGSDNSEPS 813
814 ESKLSCDPLDLFNIONNIPVMSMEDSLNSLSOLFMEIYEKEMVCMYKLDNDKIK 873
814 KNEKSCDPLDLFNIONNIPVMSMEDSLNSLSOLFMEIYEKEMVCMYKLDNDKIK 873
874 NLEBAKV-----STSVKTLSSS-----MOPSLTPQDKPEVSAND--D 912
874 KLEBOKQOITGSTSSPGTWTVTATQASATSHNSOQNSASSSTNTONGVASSGPAVE 933
913 TSHS-----TNLNSIKLFENILSGKKNKIYOL-IGOKSENEYEKILMDSPFYNESEF 967
934 ESDPLTLVLSISNDLGIYSLNDGNKTKVNPPLTJSTTEMEKFEYMLKNNDPYFNDI 993
968 TNFVSKADDINSLEDSKRRKLEEDINKLKTLOLSFDLYNNKYKLELEFDPKAKVKG 1027
994 KOVVKNSKYITGLT-ETOKNALNDEIKKLDTLOLSFDLYNNKYKLELEFDPKAKVKG 1052
1028 YKNOIKKLLILKQOESKLSLNNPKHVLFONFSVFENKKKEAEIETENTLENTKILKH 1087
1053 DKMOIKKLLILKQOESKLSLNNPKHVLFONFSVFENKKKEAEIETENTLENTKILKH 1112
1088 YKGLVYNGESSPLKTLSESTIOTEDNVASLENFKVLSLEKLDKNDLLEKKKLSYLS 1147
1113 YKGLVYNGESSPLKTLSESTIOTEDNVASLENFKVLSLEKLDKNDLLEKKKLSYLS 1172
1148 SGDLHLIALEKEVYIKNNKTYGNSPENNTDVNNALEKFKFLEPGTAVATVASE----- 1201
1173 SGDLHLIALEKEVYIKNNKTYGNSPENNTDVNNALEKFKFLEPGTAVATVASE----- 1231
1202 -----SGSDTLEOSQPKKPASTHVAGESNTITTSQNVDEVDVIVIFGESEEDY 1253

1232 PSLPVSRSKSSGSKTEETQIPGSGSLTLELOQVOSONDEEDDSLVLPFGESEEND 1291
1254 DDGQVYVGEAVTPSVADNITLSENEEYVLYKPLAGVYRSLLKOLENNVMTNVAKD 1313
1292 EYLDQVYVGEAVTPSVADNITLSENEEYVLYKPLAGVYRSLLKOLENNVMTNVAKD 1350
1314 ILNSRPNKRENFKNVLESDILIPYKDLTSSNVVADPKFLNKEKRDKFLSSYNTKSID 1373
1351 ILNSRPNKRENFKNVLESDILIPYKDLTSSNVVADPKFLNKEKRDKFLSSYNTKSID 1410
1374 TDINPANDVLGYKYLSEKYSKSDISIKKYI-----NDKOGENEKY 1414
1411 NDIFAEQEGISYEVKYLAKYKDDLESIKKYIKEEKEPPSPPTTPSPAKTDEQKESKE 1470
1415 LPLNLTETLYKYNKIDLEFLVHLEKAVLYNTYKESNVEYKIKELVLYKTIDQKLDFK 1474
1471 LPLNLTETLYKYNKIDLEFLVHLEKAVLYNTYKESNVEYKIKELVLYKTIDQKLDFK 1530
1475 KNNFVGIADSLDTPDYNNNLTKFLSTGAVENLAKTVLSNLDGNLGMNLISQHCYK 1534
1531 NTNDFEAIKLLINDYTKKMDLGLSTGLV-QNFPNTIISKLEKIQDMNLISQHCYK 1589
1535 KQCPONSCEFRHLDEREKCLLYNKQEGDKYVENPPTCNENNGGADAKTEEDSGS 1594
1590 KQCPONSCEFRHLDEREKCLLYNKQEGDKYVENPPTCNENNGGADAKTEEDSGS 1649
1595 NGKITTECTKPDSPYLPEDGIFGSSNPLGIFLILMLILYSPI 1639
1650 NGKITTECTKPDSPYLPEDGIFGSSNPLGIFLILMLILYSPI 1694

RESULT 4
025922 PRELIMINARY; PRT: 1720 AA.
AC 025922;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PRECURSOR OF THE MAJOR MEROZOITE SURFACE ANTIGENS.
OS Plasmodium falciparum (isolate NF54).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5843;
RN [1]
RP SEQUENCE OF 1669-1720 FROM N.A.
RC STRAIN=NF54;
RA Tolle R., Bujard H., Cooper J.A.:
RL Exp. Parasitol. 0:0-0(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NF54;
RA Tolle R.:
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NF54;
RX MEDLINE=96123395; PubMed=8577332;
RA Pan W., Tolle R., Bujard H.:
RT "A direct and rapid sequencing strategy for the Plasmodium falciparum
antigen gene gp190/msx".
RL Mol. Biochem. Parasitol. 73:241-244(1995).
DR EMBL: 235327; CAA84556.1; -
DR InterPro: IPR000561; -
DR Pfam: PF00008; EGF.1.
KW Signal; Merozoite.
FT SIGNAL 1 19 POTENTIAL.
SQ SEQUENCE 1720 AA; 195725 MW; 717B2FD1E637A8A3 CRC64;

Query Match 60.6%; Score 5107.5; DB 5; Length 1720;
Best Local Similarity 60.2%; Pred. No. 6; 6e-181;
Matches 1052; Conservative 216; Mismatches 343; Indels 137; Gaps 29;

```

QY 1 MKIIFFLCSPLEFIINOCVTHESYOELVKKLEADAVLTGYSLFQKEMVLANE--GT 57
D 1 MKIIFFLCSPLEFIINOCVTHESYOELVKKLEADAVLTGYSLFQKEMVLANEELTT 60
QY 58 SGTAVTSTPSSKSVASGSGSVASGSGSVASGSGSVASGSGSVASGSGN--SRTN-- 112
D 61 KGASQSGASQSG--ASAQSGASQSGASQSGASQSGSCTSPSGSTSPSSKSNLTP 118
QY 113 -----PSDNDSDAKSYADLKHVRNVLITIKELKAPQFEDLJNHNLTCDNTHG 163
D 119 RSNMSSGASPADSDSAPKSYADLKHVRNVLITIKELKAPQFEDLJNHNLTCDNTHG 178
QY 164 FKYLIDGEEINELLYKLNFFDLRAKLVANDVCANDYCOIPEPNLKRANELDVLKLVFG 223
D 179 FKYLIDGEEINELLYKLNFFDLRAKLVANDVCANDYCOIPEPNLKRANELDVLKLVFG 238
QY 224 YRKLPLNDIKDVGKMEYIKKNNKKTENINELIEBSKKTIDKNNKATKEEKKLYQAOY 283
D 239 YRKLPLNDIKDVGKMEYIKKNNKKTENINELIEBSKKTIDKNNKADNEGGKKLYQAOY 298
QY 284 DLSIYNKQLEBAHNLISLEKRIDPLKKNENIKELDKINELKPNPANSNGTPTLTDK 343
D 299 DLSIYNKQLEBAHNLISLEKRIDPLKKNENIKELDKINELKPNPANSNGTPTLTDK 358
QY 344 NKIIIEHEKEIKETAKTIFENIDSLFTDPLELEYLREKKNIDISAKVE--TKESTEPN 401
D 359 NKIIIEHEKEIKETAKTIFENIDSLFTDPLELEYLREKKNKVDTPKSOPTKVOQIPK 418
QY 402 -EYNGVYPLSYDINNAL--NELNSFGDLINFDYTKER--SKNIYTDN--ERKKFTNE 455
D 419 VPYNGVYPLPLDINHSLADNDKNSYGDLMNP--HTKEKINELITDMEKKEKIFITNN 476
QY 456 IKKIKIKIKKKI-----ESDKSYEDRSKSLNDITREYELKINELIYDSKNNNIDLTNEK 511
D 477 IKKIKIDLEKKNINHTKEONKRLDEYKS---KDYBELLEKFEKFNPNEDVDVVK 532
QY 512 MMGRYSYKVEKLTHTNHFASYSKHNLEKLTAKLYMEDYSLENIYVEKELKYYKMLI 571
D 533 IFSARYTYNVEKRYNNKFSSNNSYVNOVKTKALSLYEDLSLKGISEKDFNHYTLK 592
QY 572 SKIENEITLVENIKKDEQLPEKK--ITDENKPDKEILEVSDIVKVOQVYLLMKI 628
D 593 TGLBADIKLTKEEIKSSENKILEKNFKGLTHSANGS---LEVSDIVKVOQVYLLMKI 648
QY 629 DELKKTQILKNVLEKHNHIVPNYSKOENKOPRYLLIYKKEIDKLYPMKVESELINE 688
D 649 EDLKKIEFLKNAOLKOSIHVPNIYKPKQKPEPYLLIYKKEIDKLYPMKVESELINE 708
QY 689 KKNIKTEGQSNSEPTGECITGOATTPKQOAGSALGDSVQAOAKOAKO-----P 742
D 709 QAVLSS-----ITQPLVASETTIEDGSHSTHTLSQSETEVTEETEETEVGHTT 759
QY 743 PVVPVPEAKAQPVPAPV-----NNKTENVSKLDELEKLEYFLNTSYICHKY 791
D 760 TVTITLPTQ---PSPREKVVENSIHKSNDNSQALTKYIYKLEDEFLTKSYICHKY 816
QY 792 ILVSHSTNENKILQYKTKTEESKLSCLDPLFNQNNIPWYSKFDLSNLSOLF 851
D 817 ILVSNSSMDOKLTLEVYNLTPEEENELKCDPLDLFNQNNIPWYSKFDLSNLSOLF 876
QY 852 MEIYERKEVNCULYKLDKDKIKNLEEKV-----STSVKTLSSSS----- 893
D 877 FELYQKEKITIYLLHKLKEKNHKLLEBKQITGTSSTSPSGTIVNVAQSTHNSQNOQ 936
QY 894 MOPILAPPODKPEVSAND--DTSHS---TNLNSKLFLFNILSKGNKIYOEL-IGQ 945
D 937 SNASTTNQNGVAVSSGPAVEESHDPPLTVLISNDLKGIYSLNKGKTVPNPLTST 996
QY 946 KSSNFYEKILKSDTFNENFTFNVVSKADDINSLDESRRKKEEDINLKLTLQSLF 1005
D 997 TEMEKFEYENILKNNDIYFNNDIKQFVNSNSKVITGLT-ETQKNMLNDELKILTLQSLF 1055
QY 1006 DLYNKKYKTLERLDDKKTIVGKYMKQIKKLLLEQLESKLSLNNPKNHVLQNSVFENK 1065

```

```

D 1056 DLYNKKYKTLERLDFENKKELEGODKMQIKKLTLLKQLESKLSLNNPKNHVLQNSVFENK 1115
QY 1066 KKEAIAETENTLENTKILKHKYKLVYVYNGESSPLKTLSEESIQTODNVALENFYL 1125
D 1116 KKEAIAETENTLENTKILKHKYKLVYVYNGESSPLKTLSEESIQTODNVALENFYL 1175
QY 1126 SKLEGLKDNILNEKKKISLSSGLHNLIAELREYKRNKNTGNSPNNNDVNNALASY 1185
D 1176 SKIDGLNDNLHGLKKKLSLSSGLHNLITELREYKRNKNTGNSPNNNDVNNALASY 1235
QY 1186 KFLPEGTDAVAVSE-----SGSDPLEOSQPKPASTVGAESNTITTSQ 1231
D 1236 ENFLPE-AKVYTVVTPPPDPVTPSPLSVRSVSSGSSSTKEEFPQIPISGLIHELQOVQLO 1294
QY 1232 NVDEVDVYIIVPIFGESEEDYDGLGVVYGBATVPYSVIDNLSIKENEYEVLYKPLAG 1291
D 1295 NYDEEDDSLVLVLPJFGESEEDNDEYLDQVYVGTGALSVT--MDNILSGFENEYDVIYKPLAG 1353
QY 1292 YVRSILKQLENNVMTFNVVNDILNSRKNRENFKNVLESOLIPYKDLTSSNYVVKDPPK 1351
D 1354 YVRSILKQIEKNITFTNLNDILNSRLKRRYFLVLESOLIPYKDLTSSNYVVKDPPK 1413
QY 1352 FLNKKERDKFLSSYNYIKOSIDTIDINFANDVLYGYYKILSEYKSDLSIKRYI----- 1404
D 1414 LNSQKQWTLKSYKIKESYENDIKFAQEGISYEKVLAKYKODLESIKRYIKEKEKF 1473
QY 1405 -----NDKQENKRYLFLNNITELTYVNDKIDLYIHLKAVLNTYKS 1451
D 1474 PSSPPTPPSPAKTDEQKESKFLPEFTNIEFTLVNMLNKNKIDDYDILNKAINDCNVEDK 1533
QY 1452 NVEYKIKELNYLKTIDKLABPKKNNFVGTADISTDVNNHNLTKFLSTGVFENLKT 1511
D 1534 EAHVAKITLSDKALDIDKIDLEKPNYDEALIKKLNDDTKKMDLKLSTGLV-ONFPNT 1592
QY 1512 VLSLIDNLOGLMNLISOHQCVKQCPONSGCFRHLDERRECKCLLNYKQESDKCVENPN 1571
D 1593 IISKIEKFPQDMLNISQHCQKQCPENSGCFRHLDERRECKCLLNYKQESDKCVENPN 1652
QY 1572 PTCNENNGCCDADAKCTEEDSGSNGKKTTCCTCKDPSYPLPDGIFCSSNPLGSLFLLIL 1631
D 1653 PTCNENNGCCDADACTEEDSGSNGKKTTCCTCKDPSYPLPDGIFCSSNPLGSLFLLIL 1712
QY 1632 MLILXSF 1639
D 1713 MLILXSF 1720

```

RESULT 5

Q26194 PRELIMINARY; PRT: 1751 AA.

AC Q26194:

DT 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)

DE (PV200).

GN PV200.

OS Plasmodium vivax.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_Taxid=5855;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SAU-1.

FX MEDLINE=92158013; PubMed=1371329;

RA Gibson H.L., Tucker J.E., Kaslow D.C., Kretzli A., Collins W.E., Kiefer M.C., Bachurst I.C., Barr P.J.,

RT "Structure and expression of the gene for Pv200, a major blood-stage surface antigen of Plasmodium vivax."

RL Mol. Biochem. Parasitol. 50:325-334(1992).

DR EMBL; M75674; AAA29735.1; -

DR HSSP; P05412; IJUN.

KW Signal.

FT SIGNAL 1 19 POTENTIAL.

Db	Q91	NDNC	-ASNHNHNSVSKSPN1	-SCAKHHSTPQAEINQVGVNSPEKPEADPAQVEKPYDKHL	988
Qy	957	KDSDFPNESFTNFVW	SKADQDINSLNDESKRKKLEEDINLKLKLTQLQSFQDLYNKKYKLE		1016
Db	989	SQIDK	-VNDYFKKFLEKSEIKM	-DPIKMAALGEIEELKKKLVQVSDHYGKYLKLE	1046
Qy	1017	RLEPKKTKVAKYKMOJK	KLKLLKLEQLESKLSNINPKHYLVQNSVFENKKKEAEIAETEN		1076
Db	1047	RELKAKKIKSNKQDK	KLKLLSLNKKERONLNNPTSLKNTYTFAPFNKKRTEKKEVEN		1106
Qy	1077	TLENFKILKHKYGLV	YNGESSPLKTSIESIQTEDNYASILENFKYLSLLEGLKQNL		1136
Db	1107	TLKNTIELLKYKARAK	YVIGEPFLKTSIESQMQEDNYLNLKLEKRSADMK	-EIKRDT	1165
Qy	1137	NLEKKLSIYSSGHLHL	LAELKLVKNNKNTGNSPSENNTDVNNALLESKTKLP	-----E	1191
Db	1166	ELERSNITSYSSGLHL	DAEELINDKKRYSGDKAHNAEKKKLVQVQELIPVTSQE		1225
Qy	1192	GTDAVAVV	-----SESGSDTLEQSPKKPASTHY	-	1220
Db	1226	STSAVAVVPGAVV	PGVPTAAAGSGASGAVPPAPAAAGSASCAVPPAGCPSPATGVVP		1265
Qy	1221	-----GASNTITTSQNV	DEVDVITVPIFGESSEEDYDLQGVVGEA	-----VPPSV	1269
Db	1286	GVESASAQKRAQAO	YAEQYDVKVIALPLFGNDDGEE	-DQVTTGEASEAPETLVAPAG	1344
Qy	1270	IDNLLKSTENEEVYL	KPLAGYRSBKQOLENNVTFNANKDLNSRFNKRENFKNVL		1329
Db	1345	I	-----SDYDVVYLKPLAGMYKTKIKQLENNVAFNTITDMLDSLKKRNFLEVL		1366
Qy	1330	ESDILPKDILSSVYV	KDQYFLNKKRDLSSVNYTKSDIDNDINPANDVGLGYKL		1389
Db	1397	NSDINLPKYPSPG	EYIIKDPYKLDLLEKKKLLGSIKYIGASIDKDLATANDGVITYYKM		1456
Qy	1390	SEKYKSDL	-----DSIKKYIND	-----KQGENEYKLP	1419
Db	1457	GELYKTHLTVANEE	KVKEADIKAEEDDKIKKIGSDSTKTEQTQSWAKKAELKYLPLN		1516
Qy	1420	NIETLYTVNDKID	LEVYHLEAKVNLVYTEKSANEVKEIKELVYLTIDOKDLAPFKKNNF		1479
Db	1517	SLOEYSLVSKVWTY	TDNLKRYINNCOLEKBAELTVYKRIQDYNNKMEKELEYKK	----	1572
Qy	1480	VGIDSLSTDYNNH	NLLTKFLSTQVMEFENLAKVYLSNLLDGNLQGMNLISOHQCVYKQCPQ		1539
Db	1573	-----SEKNEVKS	SSGLEKLMKSKLILENBSKELISLQSLVQVQLTMSSEHFCIDITVNP		1628
Qy	1540	NSGCFRHLDEREC	CKLLANYKQBGDCVNPNTCENNNGCGDADAKTEEDSGSNGKI		1599
Db	1630	NAACYRFLDQME	EMRCLTFKEBGGCVGSNVTCKDNNGGCAPAEKCMTDS	-----NKI	1685
Qy	1600	TCECTKPDSTPL	PDGIFGCSSNFIIGISFLIIMLIL		1635
Db	1686	VCKCTKEGSEPL	EGVGCSSSSFLSFLIIMLFL		1721
RESULT	8				
Q9TYGI		PRELIMINARY:	PRT:	539	AA.
AC	Q9TYGI:				
DT	01-MAY-2000	(T-EMBLrel. 13, Created)			
DT	01-MAY-2000	(T-EMBLrel. 13, Last sequence update)			
DT	01-JUN-2000	(T-EMBLrel. 14, Last annotation update)			
DE	MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).				
GN	MSP1.				
OS	Plasmodium falciparum.				
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
CX	NCBI_TaxID=5833;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=93295445; PubMed=8515786;				
RA	Jongwitties S., Tanabe K., Kanbara H.;				
RT	"Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP) of Plasmodium falciparum from				

RT field isolates.
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13358; BAA02619.1; -
DR InterPro: IPR000561; -
DR InterPro: IPR001245; -
DR Pfam: PF00008; EGF; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61075 MW; C70C2E100ECA101 CRC64;

Query Match 33.3%; Score 2801; DB 5; Length 539;
Best Local Similarity 100.0%; Pred. No. 2.2e-96;
Matches 539; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1101 PLKLTSESIQTEDNYASLENFKVLKLEGLKDKNLNLEKKRLSYLSSGLHLLAELEKEV 1160
DB 1 PLKLTSESIQTEDNYASLENFKVLKLEGLKDKNLNLEKKRLSYLSSGLHLLAELEKEV 60
QY 1161 IKNNYTGNSPSENNVDVNNALLESYKFKFLPGCTDVAIVVSSGSDTLEOSQPKKPASTHV 1220
DB 61 IKNNYTGNSPSENNVDVNNALLESYKFKFLPGCTDVAIVVSSGSDTLEOSQPKKPASTHV 120
QY 1221 GAESNTITTSQNVDEVDVIVPIFGSEEDYDLDGQVTVGEAVTPSVIDNLSKIENE 1280
DB 121 GAESNTITTSQNVDEVDVIVPIFGSEEDYDLDGQVTVGEAVTPSVIDNLSKIENE 180
QY 1281 YEVLKPLAGVYRSKLENNVWFENVNKKDILNSFRNREKNVLESDLIPYKDLT 1340
DB 181 YEVLKPLAGVYRSKLENNVWFENVNKKDILNSFRNREKNVLESDLIPYKDLT 240
QY 1341 SSNVVVKDPYKFLNKEKDKFLSSYNYIKDSIDPDINFANVLYGKYLSEKYSKSDLSI 1400
DB 241 SSNVVVKDPYKFLNKEKDKFLSSYNYIKDSIDPDINFANVLYGKYLSEKYSKSDLSI 300
QY 1401 KKYINDKOGENEKYLPLNNIETLYKTVNDRKIDLFVHLEAKVLYTYEKSNEVKIKEL 1460
DB 301 KKYINDKOGENEKYLPLNNIETLYKTVNDRKIDLFVHLEAKVLYTYEKSNEVKIKEL 360
QY 1461 NYLKTIDQKLADFKKNNFVGIADLSTDYNNHNNLTKFLSGWGFENLAKTVLSMLDGN 1520
DB 361 NYLKTIDQKLADFKKNNFVGIADLSTDYNNHNNLTKFLSGWGFENLAKTVLSMLDGN 420
QY 1521 LOGMLNISOHCYKQCPONGSCFRHLDERECKCLLYKQEGDKCVENPPTCENNNG 1580
DB 421 LOGMLNISOHCYKQCPONGSCFRHLDERECKCLLYKQEGDKCVENPPTCENNNG 480
QY 1581 CDADAKCTEEDSGSNGKKTCECTKPDSPYPLFDGIFCSSNPLGISFLILMLILYSFI 1639
DB 481 CDADAKCTEEDSGSNGKKTCECTKPDSPYPLFDGIFCSSNPLGISFLILMLILYSFI 539

RESULT 9
Q25972 PRELIMINARY; PRT; 539 AA.
AC Q25972;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
DE 01-JUN-2000 (TREMblrel. 14, last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; Pubmed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).

DR EMBL; D13360; BAA02621.1; -
DR InterPro: IPR000561; -
DR InterPro: IPR001245; -
DR Pfam: PF00008; EGF; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61076 MW; 72186FF1412C57D1C CRC64;

Query Match 33.2%; Score 2798; DB 5; Length 539;
Best Local Similarity 99.8%; Pred. No. 2.8e-96;
Matches 538; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1101 PLKLTSESIQTEDNYASLENFKVLKLEGLKDKNLNLEKKRLSYLSSGLHLLAELEKEV 1160
DB 1 PLKLTSESIQTEDNYASLENFKVLKLEGLKDKNLNLEKKRLSYLSSGLHLLAELEKEV 60
QY 1161 IKNNYTGNSPSENNVDVNNALLESYKFKFLPGCTDVAIVVSSGSDTLEOSQPKKPASTHV 1220
DB 61 IKNNYTGNSPSENNVDVNNALLESYKFKFLPGCTDVAIVVSSGSDTLEOSQPKKPASTHV 120
QY 1221 GAESNTITTSQNVDEVDVIVPIFGSEEDYDLDGQVTVGEAVTPSVIDNLSKIENE 1280
DB 121 GAESNTITTSQNVDEVDVIVPIFGSEEDYDLDGQVTVGEAVTPSVIDNLSKIENE 180
QY 1281 YEVLKPLAGVYRSKLENNVWFENVNKKDILNSFRNREKNVLESDLIPYKDLT 1340
DB 181 YEVLKPLAGVYRSKLENNVWFENVNKKDILNSFRNREKNVLESDLIPYKDLT 240
QY 1341 SSNVVVKDPYKFLNKEKDKFLSSYNYIKDSIDPDINFANVLYGKYLSEKYSKSDLSI 1400
DB 241 SSNVVVKDPYKFLNKEKDKFLSSYNYIKDSIDPDINFANVLYGKYLSEKYSKSDLSI 300
QY 1401 KKYINDKOGENEKYLPLNNIETLYKTVNDRKIDLFVHLEAKVLYTYEKSNEVKIKEL 1460
DB 301 KKYINDKOGENEKYLPLNNIETLYKTVNDRKIDLFVHLEAKVLYTYEKSNEVKIKEL 360
QY 1461 NYLKTIDQKLADFKKNNFVGIADLSTDYNNHNNLTKFLSGWGFENLAKTVLSMLDGN 1520
DB 361 NYLKTIDQKLADFKKNNFVGIADLSTDYNNHNNLTKFLSGWGFENLAKTVLSMLDGN 420
QY 1521 LOGMLNISOHCYKQCPONGSCFRHLDERECKCLLYKQEGDKCVENPPTCENNNG 1580
DB 421 LOGMLNISOHCYKQCPONGSCFRHLDERECKCLLYKQEGDKCVENPPTCENNNG 480
QY 1581 CDADAKCTEEDSGSNGKKTCECTKPDSPYPLFDGIFCSSNPLGISFLILMLILYSFI 1639
DB 481 CDADAKCTEEDSGSNGKKTCECTKPDSPYPLFDGIFCSSNPLGISFLILMLILYSFI 539

RESULT 10
Q25966 PRELIMINARY; PRT; 539 AA.
AC Q25966;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
DE 01-JUN-2000 (TREMblrel. 14, last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; Pubmed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL EMBL; D13357; BAA02618.1; -
DR InterPro: IPR000561; -

DR InterPro: IPR001245; -
DR Pfam: PF00008; EGF; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 539 AA: 61144 MW: 787DE90C1D0ACDE7 CRC64;

Query Match 33.2%; Score 2796; DB 5: Length 539;
Best Local Similarity 99.8%; Pred. No. 3.3e-96;
Matches 538; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1101 PLKLTSESIQTEDNYASLENFKVLSKLEGLKDNLEKKKLSYLSGHLHLIAELKEV 1160
DB 1 PLKLTSESIQTEDNYASLENFKVLSKLEGLKDNLEKKKLSYLSGHLHLIAELKEV 60
QY 1161 IKNNKNTGNSPSENTDVNNALESYKFKFLPGTDVATVYSSGDTLQSQPKKASTHV 1220
DB 61 IKNNKNTGNSPSENTDVNNALESYKFKFLPGTDVATVYSSGDTLQSQPKKASTHV 120
QY 1221 GAESNTITTSQNVDEVDVITVPIFGSESEDDYDGLGVVTGEAVTPSVIDNIIISKIENE 1280
DB 121 GAESNTITTSQNVDEVDVITVPIFGSESEDDYDGLGVVTGEAVTPSVIDNIIISKIENE 180
QY 1281 YEVLTKPLAGYRSLSKQLENNVMTFVNWKDILNSRFNKNREKNVLESDLIPYKDLT 1340
DB 181 YEVLTKPLAGYRSLSKQLENNVMTFVNWKDILNSRFNKNREKNVLESDLIPYKDLT 240
QY 1341 SSNNVVKDPYKFLNKEKDKFLSSYNYIKDSITDIDINADVLYGKYLISEKYSDDLDSI 1400
DB 241 SSNNVVKDPYKFLNKEKDKFLSSYNYIKDSITDIDINADVLYGKYLISEKYSDDLDSI 300
QY 1401 KKYINDKOGENEKYLPLFNNIETLYKTVNDRIDLFVHLEAKVLYTYEKSNVKIKEL 1460
DB 301 KKYINDKOGENEKYLPLFNNIETLYKTVNDRIDLFVHLEAKVLYTYEKSNVKIKEL 360
QY 1461 NYLKTIOQLADFKKNNFVGIADLSTDYNNHNNLLTKFLSGMVFENLAKTVLSNLDGN 1520
DB 361 NYLKTIOQLADFKKNNFVGIADLSTDYNNHNNLLTKFLSGMVFENLAKTVLSNLDGN 420
QY 1521 LOGMNTISOHCYKQCKQNSGCFRHLDERECKCLLYKQBGKCVENPPTCENNNG 1580
DB 421 LOGMNTISOHCYKQCKQNSGCFRHLDERECKCLLYKQBGKCVENPPTCENNNG 480
QY 1581 CDADAKCTEEDSGSNGKKTCECTKPDSPLEDFGIFCSSNPLGISFLIIMLIIYSFI 1639
DB 481 CDADAKCTEEDSGSNGKKTCECTKPDSPLEDFGIFCSSNPLGISFLIIMLIIYSFI 539

RESULT 11
Q25973 PRELIMINARY: PRT: 539 AA.
AC Q25973;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
RT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwitties S.; Tanabe K.; Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13361; BAA02622.1; -
DR InterPro: IPR000561; -
DR Pfam: PF00008; EGF; 1.
KW Merozoite; EGF-like domain.

DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 539 AA: 61047 MW: 105C2E02974FB398 CRC64;

Query Match 33.2%; Score 2795; DB 5: Length 539;
Best Local Similarity 99.8%; Pred. No. 3.6e-96;
Matches 538; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1101 PLKLTSESIQTEDNYASLENFKVLSKLEGLKDNLEKKKLSYLSGHLHLIAELKEV 1160
DB 1 PLKLTSESIQTEDNYASLENFKVLSKLEGLKDNLEKKKLSYLSGHLHLIAELKEV 60
QY 1161 IKNNKNTGNSPSENTDVNNALESYKFKFLPGTDVATVYSSGDTLQSQPKKASTHV 1220
DB 61 IKNNKNTGNSPSENTDVNNALESYKFKFLPGTDVATVYSSGDTLQSQPKKASTHV 120
QY 1221 GAESNTITTSQNVDEVDVITVPIFGSESEDDYDGLGVVTGEAVTPSVIDNIIISKIENE 1280
DB 121 GAESNTITTSQNVDEVDVITVPIFGSESEDDYDGLGVVTGEAVTPSVIDNIIISKIENE 180
QY 1281 YEVLTKPLAGYRSLSKQLENNVMTFVNWKDILNSRFNKNREKNVLESDLIPYKDLT 1340
DB 181 YEVLTKPLAGYRSLSKQLENNVMTFVNWKDILNSRFNKNREKNVLESDLIPYKDLT 240
QY 1341 SSNNVVKDPYKFLNKEKDKFLSSYNYIKDSITDIDINADVLYGKYLISEKYSDDLDSI 1400
DB 241 SSNNVVKDPYKFLNKEKDKFLSSYNYIKDSITDIDINADVLYGKYLISEKYSDDLDSI 300
QY 1401 KKYINDKOGENEKYLPLFNNIETLYKTVNDRIDLFVHLEAKVLYTYEKSNVKIKEL 1460
DB 301 KKYINDKOGENEKYLPLFNNIETLYKTVNDRIDLFVHLEAKVLYTYEKSNVKIKEL 360
QY 1461 NYLKTIOQLADFKKNNFVGIADLSTDYNNHNNLLTKFLSGMVFENLAKTVLSNLDGN 1520
DB 361 NYLKTIOQLADFKKNNFVGIADLSTDYNNHNNLLTKFLSGMVFENLAKTVLSNLDGN 420
QY 1521 LOGMNTISOHCYKQCKQNSGCFRHLDERECKCLLYKQBGKCVENPPTCENNNG 1580
DB 421 LOGMNTISOHCYKQCKQNSGCFRHLDERECKCLLYKQBGKCVENPPTCENNNG 480
QY 1581 CDADAKCTEEDSGSNGKKTCECTKPDSPLEDFGIFCSSNPLGISFLIIMLIIYSFI 1639
DB 481 CDADAKCTEEDSGSNGKKTCECTKPDSPLEDFGIFCSSNPLGISFLIIMLIIYSFI 539

RESULT 12
Q25976 PRELIMINARY: PRT: 539 AA.
AC Q25976;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
RT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwitties S.; Tanabe K.; Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13362; BAA02623.1; -
DR InterPro: IPR000561; -
DR Pfam: PF00008; EGF; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW Merozoite; EGF-like domain.

FT NON_TER 1 1
SQ SEQUENCE 539 AA; 61045 MW; 8BF9C64322E9A778 CRC64;

Query Match 33.2%; Score 2794; DB 5; Length 539;
Best Local Similarity 99.8%; Pred. No. 3.9e-96;

Matches 538; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1101 PLKLTSEESIOTEDNYASLENFKVLSKLEGLKDKLNLEKKKLSYSSGLHLLIAELKEV 1160
DB 1 PLKLTSEESIOTEDNYASLENFKVLSKLEGLKDKLNLEKKKLSYSSGLHLLIAELKEV 60
QY 1161 IKNNYTGNSPSENNNTDVNNALESYKKFLPGCTDVATVVSSEGSPTLEQSPKKPASTHV 1220
DB 61 IKNNYTGNSPSENNNTDVNNALESYKKFLPGCTDVATVVSSEGSPTLEQSPKKPASTHV 120
QY 1221 GAESNTITTSQNVDEVDVIVPIFGSESEDDYDLGQVTVGEAVTPSVINDILISKIENE 1280
DB 121 GAESNTITTSQNVDEVDVIVPIFGSESEDDYDLGQVTVGEAVTPSVINDILISKIENE 180
QY 1281 YEVLTKPLAGVYSRLKKOLENNMTFNVNWKDILNSRFNKRNFKNVLESDLIPYKDLT 1340
DB 181 YEVLTKPLAGVYSRLKKOLENNMTFNVNWKDILNSRFNKRNFKNVLESDLIPYKDLT 240
QY 1341 SSNVVVKDPYKFLNKKERDKFLSSYNYIKDSIDTDINPANDVLAGYKILSEKYSDDLST 1400
DB 241 SSNVVVKDPYKFLNKKERDKFLSSYNYIKDSIDTDINPANDVLAGYKILSEKYSDDLST 300
QY 1401 KKYINDKOGENEKYLPLNNIETLYKTVNDKIDLFVHLEAKVNLVTEKSNVEVKIKEL 1460
DB 301 KKYINDKOGENEKYLPLNNIETLYKTVNDKIDLFVHLEAKVNLVTEKSNVEVKIKEL 360
QY 1461 NYLKTIDOKLADFKNNNFVGIADLSTQYNNNNLTFLSTGMPFENLAKTVLSMLDGN 1520
DB 361 NYLKTIDOKLADFKNNNFVGIADLSTQYNNNNLTFLSTGMPFENLAKTVLSMLDGN 420
QY 1521 LOGMLNISOHCYKQCPONSQCFRHLDERECKCLLNYKQEGDKCVENPPTCENNNG 1580
DB 421 LOGMLNISOHCYKQCPONSQCFRHLDERECKCLLNYKQEGDKCVENPPTCENNNG 480
QY 1581 CDADAKCTEEDSGSNGKKTCECTKPDSPYPLFDGIFGSSSNFLGISFLIIMLILYSFI 1639
DB 481 CDADAKCTEEDSGSNGKKTCECTKPDSPYPLFDGIFGSSSNFLGISFLIIMLILYSFI 539

RESULT 13
Q25981
ID Q25981 PRELIMINARY; PRT; 539 AA.
AC Q25981;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwittiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates."
RL MOL. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13363; BAA02624.1; -
DR InterPro: IPR000561; -
DR InterPro: IPR001245; -
DR Pfam: PF00008; EGF.1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1 1
SQ SEQUENCE 539 AA; 61046 MW; 3EBDB7473EB87B65 CRC64;

Query Match 33.1%; Score 2791; DB 5; Length 539;
Best Local Similarity 99.6%; Pred. No. 5.1e-96;
Matches 537; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1101 PLKLTSEESIOTEDNYASLENFKVLSKLEGLKDKLNLEKKKLSYSSGLHLLIAELKEV 1160
DB 1 PLKLTSEESIOTEDNYASLENFKVLSKLEGLKDKLNLEKKKLSYSSGLHLLIAELKEV 60
QY 1161 IKNNYTGNSPSENNNTDVNNALESYKKFLPGCTDVATVVSSEGSPTLEQSPKKPASTHV 1220
DB 61 IKNNYTGNSPSENNNTDVNNALESYKKFLPGCTDVATVVSSEGSPTLEQSPKKPASTHV 120
QY 1221 GAESNTITTSQNVDEVDVIVPIFGSESEDDYDLGQVTVGEAVTPSVINDILISKIENE 1280
DB 121 GAESNTITTSQNVDEVDVIVPIFGSESEDDYDLGQVTVGEAVTPSVINDILISKIENE 180
QY 1281 YEVLTKPLAGVYSRLKKOLENNMTFNVNWKDILNSRFNKRNFKNVLESDLIPYKDLT 1340
DB 181 YEVLTKPLAGVYSRLKKOLENNMTFNVNWKDILNSRFNKRNFKNVLESDLIPYKDLT 240
QY 1341 SSNVVVKDPYKFLNKKERDKFLSSYNYIKDSIDTDINPANDVLAGYKILSEKYSDDLST 1400
DB 241 SSNVVVKDPYKFLNKKERDKFLSSYNYIKDSIDTDINPANDVLAGYKILSEKYSDDLST 300
QY 1401 KKYINDKOGENEKYLPLNNIETLYKTVNDKIDLFVHLEAKVNLVTEKSNVEVKIKEL 1460
DB 301 KKYINDKOGENEKYLPLNNIETLYKTVNDKIDLFVHLEAKVNLVTEKSNVEVKIKEL 360
QY 1461 NYLKTIDOKLADFKNNNFVGIADLSTQYNNNNLTFLSTGMPFENLAKTVLSMLDGN 1520
DB 361 NYLKTIDOKLADFKNNNFVGIADLSTQYNNNNLTFLSTGMPFENLAKTVLSMLDGN 420
QY 1521 LOGMLNISOHCYKQCPONSQCFRHLDERECKCLLNYKQEGDKCVENPPTCENNNG 1580
DB 421 LOGMLNISOHCYKQCPONSQCFRHLDERECKCLLNYKQEGDKCVENPPTCENNNG 480
QY 1581 CDADAKCTEEDSGSNGKKTCECTKPDSPYPLFDGIFGSSSNFLGISFLIIMLILYSFI 1639
DB 481 CDADAKCTEEDSGSNGKKTCECTKPDSPYPLFDGIFGSSSNFLGISFLIIMLILYSFI 539

RESULT 14
Q25984
ID Q25984 PRELIMINARY; PRT; 539 AA.
AC Q25984;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwittiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates."
RL MOL. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13356; BAA02617.1; -
DR InterPro: IPR000561; -
DR InterPro: IPR001245; -
DR Pfam: PF00008; EGF.1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1 1
SQ SEQUENCE 539 AA; 61114 MW; 3788015F3127C9E CRC64;

Query Match 33.1%; Score 2789; DB 5; Length 539;
 Best Local Similarity 99.6%; Pred. No. 6e-96;
 Matches 537; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1101 PLKLTSEESIOTEDNYASLENFKVLSKLEGLKDKONLNEKKRLSYSSGLHHLIAELKEV 1160
 |||||||
 DB 1 PLKLTSEESIOTEDNYASLENFKVLSKLEGLKDKONLNEKKRLSYSSGLHHLIAELKEV 60
 |||||||

QY 1161 IKNNYTGNSPSENNITDVNNALLESYKKFLPEGTDAVAVSSGSDTLEQSPKKPASTHV 1220
 |||||||
 DB 61 IKNNYTGNSPVNNTDVNNALLESYKKFLPEGTDAVAVSSGSDTLEQSPKKPASTHV 120
 |||||||

QY 1221 GAESNTTTSQNVDEVDVITVPIFGSEEDYDLDGQVYTGAVTPSVIDNITLSKIENE 1280
 |||||||
 DB 121 GAESNTTTSQNVDEVDVITVPIFGSEEDYDLDGQVYTGAVTPSVIDNITLSKIENE 180
 |||||||

QY 1281 YEVLTKPLAGVYRSLSKKOLENNVMTFNVKDILNSPKNREPNKNVLESDLIPYKDLT 1340
 |||||||
 DB 181 YEVLTKPLAGVYRSLSKKOLENNVMTFNVKDILNSPKNREPNKNVLESDLIPYKDLT 240
 |||||||

QY 1341 SSNVVVDVDPYKFLNKKERKDFLSSYNYIKDSIDPDINFANDVLYGYYKILSEKYSDDL 1400
 |||||||
 DB 241 SSNVVVDVDPYKFLNKKERKDFLSSYNYIKDSIDPDINFANDVLYGYYKILSEKYSDDL 300
 |||||||

QY 1401 KKYINDKOGENEKYLPELNNIETLYKTVNDKIDLFVHLEAKVLYTYEKSNEVEKIKEL 1460
 |||||||
 DB 301 KKYINDKOGENEKYLPELNNIETLYKTVNDKIDLFVHLEAKVLYTYEKSNEVEKIKEL 360
 |||||||

QY 1461 NYLKTIOOKLADFKKNNNFVGIADLSTDYNNHNLTKFLSTGMPENLAKTVLSNLDGN 1520
 |||||||
 DB 361 NYLKTIOOKLADFKKNNNFVGIADLSTDYNNHNLTKFLSTGMPENLAKTVLSNLDGN 420
 |||||||

QY 1521 LOGMLNISOHCYKCKOCQNSGCFRHLDERECKCLLNYKQEGDKCVENPPTCENNNG 1580
 |||||||
 DB 421 LOGMLNISOHCYKCKOCQNSGCFRHLDERECKCLLNYKQEGDKCVENPPTCENNNG 480
 |||||||

QY 1581 CDADAKCTEEDSGSNGKKITCECTKPDSPPLFDGIFGSSNPLGISFLILMLILYSPI 1639
 |||||||
 DB 481 CDADAKCTEEDSGSNGKKITCECTKPDSPPLFDGIFGSSNPLGISFLILMLILYSPI 539
 |||||||

RESULT 15
 Q25971 PRELIMINARY; PRT: 539 AA.

ID Q25971
 AC Q25971;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
 DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
 GN MSP1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9329445; PubMed=8515786;
 RA Jongwutives S., Tanabe K., Kanbara H.;
 RT "Sequence conservation in the C-terminal part of the precursor to the
 RT major merozoite surface proteins (MSP) of Plasmodium falciparum from
 RT field isolates";
 RL Mol. Biochem. Parasitol. 59:95-100(1993).
 DR EMBL; D13359; BAA02620.1; -
 DR InterPro; IPR000561; -
 DR InterPro; IPR001245; -
 DR Pfam; PF00008; EGF.1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR. 1.
 KW Merozoite; EGF-like domain.
 FT NON_TER 1
 SQ SEQUENCE 539 AA; 61017 MW; SCA9C651BB62B5EI CRC64;

Matches 537; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1101 PLKLTSEESIOTEDNYASLENFKVLSKLEGLKDKONLNEKKRLSYSSGLHHLIAELKEV 1160
 |||||||
 DB 1 PLKLTSEESIOTEDNYASLENFKVLSKLEGLKDKONLNEKKRLSYSSGLHHLIAELKEV 60
 |||||||

QY 1161 IKNNYTGNSPSENNITDVNNALLESYKKFLPEGTDAVAVSSGSDTLEQSPKKPASTHV 1220
 |||||||
 DB 61 IKNNYTGNSPVNNTDVNNALLESYKKFLPEGTDAVAVSSGSDTLEQSPKKPASTHV 120
 |||||||

QY 1221 GAESNTTTSQNVDEVDVITVPIFGSEEDYDLDGQVYTGAVTPSVIDNITLSKIENE 1280
 |||||||
 DB 121 GAESNTTTSQNVDEVDVITVPIFGSEEDYDLDGQVYTGAVTPSVIDNITLSKIENE 180
 |||||||

QY 1281 YEVLTKPLAGVYRSLSKKOLENNVMTFNVKDILNSPKNREPNKNVLESDLIPYKDLT 1340
 |||||||
 DB 181 YEVLTKPLAGVYRSLSKKOLENNVMTFNVKDILNSPKNREPNKNVLESDLIPYKDLT 240
 |||||||

QY 1341 SSNVVVDVDPYKFLNKKERKDFLSSYNYIKDSIDPDINFANDVLYGYYKILSEKYSDDL 1400
 |||||||
 DB 241 SSNVVVDVDPYKFLNKKERKDFLSSYNYIKDSIDPDINFANDVLYGYYKILSEKYSDDL 300
 |||||||

QY 1401 KKYINDKOGENEKYLPELNNIETLYKTVNDKIDLFVHLEAKVLYTYEKSNEVEKIKEL 1460
 |||||||
 DB 301 KKYINDKOGENEKYLPELNNIETLYKTVNDKIDLFVHLEAKVLYTYEKSNEVEKIKEL 360
 |||||||

QY 1461 NYLKTIOOKLADFKKNNNFVGIADLSTDYNNHNLTKFLSTGMPENLAKTVLSNLDGN 1520
 |||||||
 DB 361 NYLKTIOOKLADFKKNNNFVGIADLSTDYNNHNLTKFLSTGMPENLAKTVLSNLDGN 420
 |||||||

QY 1521 LOGMLNISOHCYKCKOCQNSGCFRHLDERECKCLLNYKQEGDKCVENPPTCENNNG 1580
 |||||||
 DB 421 LOGMLNISOHCYKCKOCQNSGCFRHLDERECKCLLNYKQEGDKCVENPPTCENNNG 480
 |||||||

QY 1581 CDADAKCTEEDSGSNGKKITCECTKPDSPPLFDGIFGSSNPLGISFLILMLILYSPI 1639
 |||||||
 DB 481 CDADAKCTEEDSGSNGKKITCECTKPDSPPLFDGIFGSSNPLGISFLILMLILYSPI 539
 |||||||

Search completed: October 27, 2001, 15:55:53
 Job time: 273 sec

Query Match 33.1%; Score 2788; DB 5; Length 539;
 Best Local Similarity 99.6%; Pred. No. 6.5e-96;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2001, 15:53:10 ; Search time 23.62 Seconds

(without alignments)
2377.000 Million cell updates/sec

Title: US-09-269-874-3

Perfect score: 8424

Sequence: 1 MKRIIFLCSEFLFIINTQCV.....SNFLGISFLILMLILYSFI 1639

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8424	100.0	1639	1 MSP1_PLAFW	P04933 plasmodium
2	8144.5	96.7	1630	1 MSP1_PLAFK	P04932 plasmodium
3	5136.5	61.0	1726	1 MSP1_PLAFB	P04934 plasmodium
4	5127.5	60.9	1726	1 MSP1_PLAFB	P50495 plasmodium
5	5069	60.2	1701	1 MSP1_PLAFM	P08569 plasmodium
6	5065	60.1	1701	1 MSP1_PLAFM	P13819 plasmodium
7	4854.5	57.6	1682	1 MSP1_PLAF3	P19598 plasmodium
8	2548.5	30.3	1772	1 MSP1_PLAF3	P13828 plasmodium
9	1187	14.1	233	1 MSP1_PLAFD	P13827 plasmodium
10	1096.5	13.0	281	1 MSP1_PLAFN	P13820 plasmodium
11	501.5	6.0	1957	1 YD86_SCHPO	Q10411 schistosomch
12	486	5.8	1790	1 USO1_YEAST	P23386 saccharomyc
13	477	5.7	2869	1 RBP1_PLAFB	Q00798 plasmodium
14	474.5	5.6	1875	1 MLP1_YEAST	Q02455 saccharomyc
15	457.5	5.4	1251	1 RBP2_PLAFB	Q00799 plasmodium
16	443.5	5.3	2663	1 CENE_HUMAN	Q02224 homo sapien
17	440	5.2	1805	1 HMM2_MYGE	P47460 mycoplasma
18	424	5.0	2230	1 GOG4_HUMAN	Q13439 homo sapien
19	423	5.0	1679	1 YIO9_YEAST	P40457 saccharomyc
20	421.5	5.0	1928	1 MYS1_YEAST	P08964 saccharomyc
21	405	4.8	2022	1 ANTL1_ONCVO	P21249 onchocerca
22	404	4.8	2116	1 MYS2_DICDI	P08799 dictyosteli
23	393	4.7	1312	1 RA50_YEAST	P12753 saccharomyc
24	386.5	4.6	2748	1 NUM1_YEAST	Q00402 saccharomyc
25	378	4.5	1818	1 HMM2_MYCPN	P75471 mycoplasma
26	361	4.3	3660	1 DMD_CHICK	P15533 gallus galli
27	358.5	4.3	2349	1 TPR_HUMAN	P12270 homo sapien
28	355.5	4.2	3210	1 CENE_HUMAN	P49454 homo sapien
29	351.5	4.2	1935	1 MYSB_RAT	P02564 rattus norv
30	349.5	4.1	1935	1 MYSB_HUMAN	P12883 homo sapien
31	349.5	4.1	2198	1 YLJ2_CABEL	P34367 caenorhabdi
32	349	4.1	1939	1 MYSB_MESAU	P13539 mesocricetu
33	348	4.1	1940	1 MYSB_CHICK	P02565 gallus galli

34	347	4.1	1225	1 SMCI_YEAST	P32908 saccharomyc
35	345.5	4.1	2649	1 BPA1_HUMAN	Q03001 homo sapien
36	344.5	4.1	1934	1 MYSB_MESAU	P13540 mesocricetu
37	344	4.1	2136	1 YCF2_MARPO	P09975 marcantia
38	343	4.1	1939	1 MYSB_HUMAN	P13533 homo sapien
39	341.5	4.1	1938	1 MYSB_MOUSE	Q02566 mus musculu
40	341.5	4.1	1978	1 MYSB_CHICK	P10587 gallus galli
41	338	4.0	1169	1 EX5B_BORBU	O51578 borrelia bu
42	338	4.0	1938	1 MYSB_RAT	P02563 rattus norv
43	337.5	4.0	1539	1 Y373_HUMAN	O15078 homo sapien
44	336.5	4.0	1427	1 REST_HUMAN	P30622 homo sapien
45	335	4.0	3685	1 DMD_HUMAN	P11532 homo sapien

ALIGNMENTS

RESULT 1					
ID	MSP1_PLAFW	STANDARD:	PRT: 1639 AA.		
AC	P04933:				
DT	13-AUG-1987 (Rel. 05, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	01-FEB-1996 (Rel. 33, Last annotation update)				
DE	MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)				
DE	(PMMSA) (P195).				
GN	MSP-1.				
OS	Plasmodium falciparum (isolate Wellcome).				
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
OX	NCBI_TaxID=5848;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=86014355; PubMed=2995820;				
RA	Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,				
RA	Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,				
RT	Freeman R.R.;				
RT	"Primary structure of the precursor to the three major surface				
RT	antigens of Plasmodium falciparum merozoites.",				
RL	Nature 317:270-273(1985).				
RN	[2]				
RP	REVIEWS.				
RA	Holder A.A.;				
RL	Submitted (MAR-1991) to the EMBL/Genbank/DBJ databases.				
CC	-I- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR				
CC	(POTENTIAL).				
CC	-I- PFM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42				
CC	KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF				
CC	MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
CC	EMBL: X02919; CAA26676.1; -				
DR	PIR: A24594; A24594.				
DR	InterPro: IPR000561; -				
DR	Pfam: PF00008; BGF; 1.				
KW	Maternal; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;				
KW	Transmembrane; GPI-anchor.				
FT	SIGNAL	1	19	POTENTIAL.	
FT	CHAIN	20	1639	MEROZOITE SURFACE PROTEIN 1.	
FT	CARBOHYD	116	116	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	268	268	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	764	764	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	768	768	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	783	783	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	844	844	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	920	920	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	964	964	N-LINKED (GLCNAC. . .) (POTENTIAL).	

FT CARBOHYD 1058 1058 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1445 1445 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1526 1526 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 1639 AA: 187618 MM: 2C235B616C87F6E CRC64:

Query Match 100.0%; Score 8424; DB 1: Length 1639;
 Best Local Similarity 100.0%; Pred. No. 3.7e-255;
 Matches 1639; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KRIIFELCSFLEFIINTOCVTHESIYQELVKKLEALEDAVLTFGYSEFQEKMYLNGTSGT 60
 DB 1 KRIIFELCSFLEFIINTOCVTHESIYQELVKKLEALEDAVLTFGYSEFQEKMYLNGTSGT 60
 QY 61 AVTSTPGSKGSVAGSGGSGSVAGSGSVAGSGSVAGSGSGGSGNRRNPSPNSSDS 120
 DB 61 AVTSTPGSKGSVAGSGGSGSVAGSGSVAGSGSVAGSGSGGSGNRRNPSPNSSDS 120
 QY 121 DAKSYADLKHVRNYLTITKELKYPQLFDLTNHHMLTCDNTHGFYLLDGYEINELLYK 180
 DB 121 DAKSYADLKHVRNYLTITKELKYPQLFDLTNHHMLTCDNTHGFYLLDGYEINELLYK 180
 QY 181 LNFYFDLLRAKLVNVCANDYCOIPENLIRANELDVLKLVGYRKLPLDNIKDNYKMD 240
 DB 181 LNFYFDLLRAKLVNVCANDYCOIPENLIRANELDVLKLVGYRKLPLDNIKDNYKMD 240
 QY 241 YIKKKKTIEMINELIESKTIKNNKATKEEKKKIYQAOYDSTYNKOLEPHNLIS 300
 DB 241 YIKKKKTIEMINELIESKTIKNNKATKEEKKKIYQAOYDSTYNKOLEPHNLIS 300
 QY 301 VLEKRIDLKNNENIKELDKINEIKNPPANGSNTPTLLDKNNKIEHEKEIKIAT 360
 DB 301 VLEKRIDLKNNENIKELDKINEIKNPPANGSNTPTLLDKNNKIEHEKEIKIAT 360
 QY 361 IKFNIDSLFTDPLELEYLREKNNKIDISAKVETESTEPNEPYGVTYPLSYNDINNAL 420
 DB 361 IKFNIDSLFTDPLELEYLREKNNKIDISAKVETESTEPNEPYGVTYPLSYNDINNAL 420
 QY 421 NELMSGGLMPDPTKRPKNITVDNERKKEINIKKIKIEKKISDKKSYDRSKS 480
 DB 421 NELMSGGLMPDPTKRPKNITVDNERKKEINIKKIKIEKKISDKKSYDRSKS 480
 QY 481 LNDITKEYEKLLNEIYDSEKFNNDIDLTNEFKMMGRSYKYEKLHHTFASYSKSNL 540
 DB 481 LNDITKEYEKLLNEIYDSEKFNNDIDLTNEFKMMGRSYKYEKLHHTFASYSKSNL 540
 QY 541 EKLTKALKYMEDYSLRNIVEREKLYKNLISKIENIEJTVENIKKDEQLFEKKITKD 600
 DB 541 EKLTKALKYMEDYSLRNIVEREKLYKNLISKIENIEJTVENIKKDEQLFEKKITKD 600
 QY 601 ENKDEKILVESDIYKOVQVLLMNKIDELKKTOLLKNNELKNIVPNSYKQENKOE 660
 DB 601 ENKDEKILVESDIYKOVQVLLMNKIDELKKTOLLKNNELKNIVPNSYKQENKOE 660
 QY 661 PYYLILVLAKEIDKLKVPMPKVESLINEKKNIKTEGOSDNPSPEGETTGATKPPGO 720
 DB 661 PYYLILVLAKEIDKLKVPMPKVESLINEKKNIKTEGOSDNPSPEGETTGATKPPGO 720
 QY 721 AGSALGSDVGAQAOEQKQAPVPVPVPEAKAQPPTPAPVNNKTENVSKIDYLEKYE 780
 DB 721 AGSALGSDVGAQAOEQKQAPVPVPVPEAKAQPPTPAPVNNKTENVSKIDYLEKYE 780
 QY 781 FLNLSYICHTIILVSHSTMNEKILKQYKITEEESKSSCPDLLEFNIONNIPMYSMF 840
 DB 781 FLNLSYICHTIILVSHSTMNEKILKQYKITEEESKSSCPDLLEFNIONNIPMYSMF 840
 QY 841 DSLNNSLSQLEMEIYERKEMVCLNKLKNDKIKNLLEAKKRVSTVKTLSSSMOPSLT 900
 DB 841 DSLNNSLSQLEMEIYERKEMVCLNKLKNDKIKNLLEAKKRVSTVKTLSSSMOPSLT 900
 QY 901 PODKREVSANDTSHSTMLNNSLKLFEENILSLGKNKNIYQELIGKSSSENYEKLKDS 960

DB 901 PODKREVSANDTSHSTMLNNSLKLFEENILSLGKNKNIYQELIGKSSSENYEKLKDS 960
 QY 961 TFYNSEFNPFKSRADDTINSINDESKRKLLEDINKLKTOLSPDLANKKIKLERFD 1020
 DB 961 TFYNSEFNPFKSRADDTINSINDESKRKLLEDINKLKTOLSPDLANKKIKLERFD 1020
 QY 1021 KKTIVGKRYKMOIKKLTLLKEOLESKLNSLNPKHVLQNEVSFPNKKKAEIETENTLEN 1080
 DB 1021 KKTIVGKRYKMOIKKLTLLKEOLESKLNSLNPKHVLQNEVSFPNKKKAEIETENTLEN 1080
 QY 1081 TKILKHKYGLVKKYNGSSPLKLTISESIOTEDNYSLENFKVLSKLEGLKDNLEK 1140
 DB 1081 TKILKHKYGLVKKYNGSSPLKLTISESIOTEDNYSLENFKVLSKLEGLKDNLEK 1140
 QY 1141 KKLSTLSSGLHLLIAELKEVKNKNYTGNSSENNTDVNNLSEYKKPLPGTDVATVVS 1200
 DB 1141 KKLSTLSSGLHLLIAELKEVKNKNYTGNSSENNTDVNNLSEYKKPLPGTDVATVVS 1200
 QY 1201 EGSDFLEQSOPKRPASTHVAESNTITTSQNVDEVDVITVPFGESEEDYDLQGV 1260
 DB 1201 EGSDFLEQSOPKRPASTHVAESNTITTSQNVDEVDVITVPFGESEEDYDLQGV 1260
 QY 1261 TGEAVTPSVIDNIIISKIENEYEVLYKPLAGVNSLKKOLENNVTFVNNKDLINSRFN 1320
 DB 1261 TGEAVTPSVIDNIIISKIENEYEVLYKPLAGVNSLKKOLENNVTFVNNKDLINSRFN 1320
 QY 1321 KRENKKNVLESDDLIPYKDLTSSNNVYKDPYKFLNKKERDKFLSSNYTKDSIDDDINAN 1380
 DB 1321 KRENKKNVLESDDLIPYKDLTSSNNVYKDPYKFLNKKERDKFLSSNYTKDSIDDDINAN 1380
 QY 1381 DVLGYKILSEKYSKSDLSIKKIYINDKOGENEKYLPLNNIETLYKYVNDKIDLFVHLE 1440
 DB 1381 DVLGYKILSEKYSKSDLSIKKIYINDKOGENEKYLPLNNIETLYKYVNDKIDLFVHLE 1440
 QY 1441 AKVLTYYEKSNEVKIKELMYLKTIDPKLADFKKNNPVGIADLSTDYNNHNLTKFLS 1500
 DB 1441 AKVLTYYEKSNEVKIKELMYLKTIDPKLADFKKNNPVGIADLSTDYNNHNLTKFLS 1500
 QY 1501 TGMVFENLAKTVLSNLDLGNLGNLISOHOCVKKOCPONSGCFPHLDERECCOLNLYK 1560
 DB 1501 TGMVFENLAKTVLSNLDLGNLGNLISOHOCVKKOCPONSGCFPHLDERECCOLNLYK 1560
 QY 1561 QEGDKCVENPPTCENNNGGDADAKCTEEDSGSKKITCECTKPDSPPLFDGIFCSSS 1620
 DB 1561 QEGDKCVENPPTCENNNGGDADAKCTEEDSGSKKITCECTKPDSPPLFDGIFCSSS 1620
 QY 1621 NFLGISFLIIMLILYSFI 1639
 DB 1621 NFLGISFLIIMLILYSFI 1639

RESULT 2
 MSP1_PLAFK STANDARD: PRT: 1630 AA.
 ID MSP1_PLAFK P04932;
 AC P04932;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MEMOZOITE SURFACE PROTEIN 1 PRECURSOR (MEMOZOITE SURFACE ANTIGENS)
 DE (PMWSA) (P190).
 GN MSP-1.
 OS Plasmodium falciparum (isolate KI / Thailand).
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=3839;
 RX SEQUENCE FROM N.A.
 RX MEDLINE=86136024; Pubmed=3004972;
 RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
 RA Stuenkelberg H., Bujard H.;
 RT "Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merozoites: studies at the genetic level.";
 RL EMBO J. 4:3823-3829(1985).

[2]
RN REVISIONS, SEQUENCE FROM N.A.
RP Pan W., Toile R., Bujard H.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -1- PPM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb.sib.ch).
CC -----
DR EMBL; X03371; CAA27070.1;
DR PIR; A25120; SAZOK1.
DR InterPro: IPR000561;
DR Pfam; PF00008; EGF; 1.
DR Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1630 MEROZOITE SURFACE PROTEIN 1.
FT DOMAIN 67 84 TRIPEPTIDE SG(TP) REPEAT.
FT TRANSMEM 1614 1630 MEMBRANE ANCHOR.
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 755 755 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 759 759 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 774 774 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 835 835 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 911 911 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 955 955 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1049 1049 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1156 1156 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1436 1436 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1517 1517 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1630 AA: 187289 MW; ADBDEC3CEB0A46322 CRC64;
Query Match 96.7%; Score 8144.5; DR 1; Length 1630;
Best Local Similarity 97.1%; Pred. No. 1.8e-246;
Matches 1591; Conservative 11; Mismatches 28; Indels 9; Gaps 2;

|||||
Db 352 IKFNIDSLFTDPLELEYLRKNNKIDISAKVETKESTPEPEYNGVYPLSYNDINNAL 411
Qy 421 NELNSFGDLINFDYTKREPSKNITYDNEKKFINKIKIEKKIESDKKSYDRSKS 480
Db 412 NELNSFGDLINFDYTKREPSKNITYDNEKKFINKIKIEKKIESDKKSYDRSKS 471
Qy 481 LNDIKKEYEKLINLEYDSKFPNNNDLTNFKEMGKRYSYKVEKLTHTHTFASYENSKNL 540
Db 472 LNDITKEYEKLINLEYDSKFPNNNDLTNFKEMGKRYSYKVEKLTHTHTFASYENSKNL 531
Qy 541 EKLTKALYMEDYSLRNIVVEKEKLYKYNLISKIENETETLENIKDEQLFEKKIKD 600
Db 532 EKLTKALYMEDYSLRNIVVEKEKLYKYNLISKIENETETLENIKDEQLFEKKIKD 591
Qy 601 ENKPEDEKILEVSDIVKVOYQVLLMANKIDELKKTQILKNVELKHNHVPSYKQENKOE 660
Db 592 ENKPEDEKILEVSDIVKVOYQVLLMANKIDELKKTQILKNVELKHNHVPSYKQENKOE 651
Qy 661 PYLLIVLTKKEIDKLKVPMPKVESLINEKKNKITPEGSDNSEPTEGBITQOATTKPGQO 720
Db 652 PYLLIVLTKKEIDKLKVPMPKVESLINEKKNKITPEGSDNSEPTEGBITQOATTKPGQO 711
Qy 721 AGSALGDSVOAQAOEQKQAPVPVPEAKAQVPTPPAPVNNKTEVNSKIDYLEKLYE 780
Db 712 AGSALGDSVOAQAOEQKQAPVPVPEAKAQVPTPPAPVNNKTEVNSKIDYLEKLYE 771
Qy 781 PLNTSYICHKYLIVSHSTNNEKILKOYKIKTEEESKSSCDPLDLFNINNNIPVMSMF 840
Db 772 PLNTSYICHKYLIVSHSTNNEKILKOYKIKTEEESKSSCDPLDLFNINNNIPVMSMF 831
Qy 841 DSLNNSLSQLEMEIYERKEMCNLYKIKNDKIKNLLEAKKAVSYSVKTLSSSSWQPLSLT 900
Db 832 DSLNNSLSQLEMEIYERKEMCNLYKIKNDKIKNLLEAKKAVSYSVKTLSSSSWQPLSLT 891
Qy 901 PODKREVSANDTSHSTMNNSLSKLFENILSLGKKNKINYOELIGOSSSENYEKLTKSD 960
Db 892 PODKREVSANDTSHSTMNNSLSKLFENILSLGKKNKINYOELIGOSSSENYEKLTKSD 951
Qy 961 TPYNESFTPNVKSADINSINDESKRRKLEEDINKLKTQLSFDLYNKKYKLERLFD 1020
Db 952 TPYNESFTPNVKSADINSINDESKRRKLEEDINKLKTQLSFDLYNKKYKLERLFD 1011
Qy 1021 KKTIVGKRYKMOIKKTLTKLEQLESKLNLSNPKHYLQNFVSFNNKKRAELAEFTENTLEN 1080
Db 1012 KKTIVGKRYKMOIKKTLTKLEQLESKLNLSNPKHYLQNFVSFNNKKRAELAEFTENTLEN 1071
Qy 1081 TKILKHYKGLVKKYNGGSSPLKTLEESIIOTEDNYASLENPKVLSKLEGLKIDNLNLEK 1140
Db 1072 TKILKHYKGLVKKYNGGSSPLKTLEESIIOTEDNYASLENPKVLSKLEGLKIDNLNLEK 1131
Qy 1141 KKLSTYSSGHLHLIAELKEVIRKNNKYTGNSPENNVDVNNALLESYKFLPGCTGVAATVS 1200
Db 1132 KKLSTYSSGHLHLIAELKEVIRKNNKYTGNSPENNVDVNNALLESYKFLPGCTGVAATVS 1191
Qy 1201 EGGSDTLEQSQPKKPASTHVAESNTITTSQNVDEVVDVITVPIFGSEEDYDGLGVY 1260
Db 1192 EGGSDTLEQSQPKKPASTHVAESNTITTSQNVDEVVDVITVPIFGSEEDYDGLGVY 1251
Qy 1261 TGEAVTPSVINILISKIENEEVLYLKLPLAGVYSLSKQLENNNWTFWANAKDIJNSRFN 1320
Db 1252 TGEAVTPSVINILISKIENEEVLYLKLPLAGVYSLSKQLENNNWTFWANAKDIJNSRFN 1311
Qy 1321 KRENFKNVLESDDLIPYKDLTSSNYYVVKPYFLNKKERDKFLSSYNTIKDSIDFDINAFN 1380
Db 1312 KRENFKNVLESDDLIPYKDLTSSNYYVVKPYFLNKKERDKFLSSYNTIKDSIDFDINAFN 1371
Qy 1381 DVLGYYKILSKRYKSDLSIKKYYINDKQENBEKYLPLNNIETLYKYVNDKIDLFVHLE 1440
Db 1372 DVLGYYKILSKRYKSDLSIKKYYINDKQENBEKYLPLNNIETLYKYVNDKIDLFVHLE 1431
Qy 1441 AKVINTYTERKSNVEKIKELMYLKTIDOKLADFKKNNFVGIDASTDYNNNNLTFLS 1500
|||||

QY	332	NSGTPMTLLDKNKKIEHEHEKEIKETAKTIFKNPISDLFPDPLELEYLREKKNIDISK	391
Db	359	NSGTPMTLLDKNKKIEHEHEKEIKETAKTIFKNPISDLFPDPLELEYLREKKNKADVPK	418
QY	392	VE--TKSTEPN--EYRPGVIVPLSYNDINNAL---NELMSFGDLINPBYTPEKSNITYT	445
Db	419	SQDPTKVSQJLPKVPYRPGIYVPLPLDINHSLAODKNSYDGLMNP--DTRKKNIEKIT	477
QY	446	DN--ERKFFINEIKERIKIEKKKI---ESDKSYEDRSKSLNDITKEKELNEYDSKF	500
Db	478	DNKERKTFINNKKQIDOLEKKINTHTEQONKKLEBYEKS---KKDYELLEKEFEYEMKF	533
QY	501	NNNIDLNFEMKMGKRSXYVEKTLHHNTPASJENSKHLEKLTALAKMDEYSRLNIVV	560
Db	534	NNNDDKDVQVCKIFESARTIVYVEKORINNNKFSSSNNNSYVAVQKIKALSLDEYSLRKGIS	593
QY	561	EKELEYKKNLISKIENLEIFLVENIKKDEOLESEK---ITKDEMKPDKLIEVSDIYKV	617
Db	594	EKDPNNHYTLKTGELAIKKTLEPESSEKILEKFKGLITSANAS---LEVADIYKL	649
QY	618	QOVQVFLMMKIDELKQTÖLLKVNVEKHNHIVPNYSKOENKQEPYLLVLVKKEDKLAVF	677
Db	650	QOVQVLLTIKIEDLRKTEFLNAQÖLKDSITHVNIYKPOÑKPEPYLLYLKREVDKLEF	709
QY	678	MPKYESLINEKKNIKKEGSDNSSESTBEITIGQATTFRGQAGALSBDVSQAQAOBQ	737
Db	710	IPYKMDIKKEQAVLSS-----ITÖPLVASETTEDGGSHSTHSLQSGEVEYTEET	760
QY	738	KQAPRPV-----PVPVEAKAOVTPPAPRNKNTENVSKLDYELEPTEFNTSYCH	789
Db	761	EEPEYVGHITTYITILPPEKVEYVENSIEHKSDNSQALTKYILKIDELTSLSYCH	820
QY	790	KYLLVSHSTINEKILKOYKITKEESKLSCDPLDLFNIONNIPVMSFMDLNNSSISQ	849
Db	821	KYLLVSSNSMOQKLEVVYNTLPPEENELKSCDPLDLFNIONNIPAMYSLYDSMNNDÖH	880
QY	850	LFMBITYEKVACNLYKIKDNDKIKNLEAKKY-----SISYKTLSSS-----	893
Db	881	LFEPLOYEKEMLYJHKKIKENHNIKILBEOKÖTGTSTSSPBNPTVNAQATNSNÖN	940
QY	894	--MÖPLSLTPDCKREVSANP---DTSHS---TNLNNSLKLEFNILSLGKNKNIYQEL-I	943
Db	941	QÖSNASSTNÖNGVAVSSGAVVEESHDPLVLVISTENDKGLVSLNLGKNKTKVNPPLTI	1000
QY	944	GÖKSSENFYEKILKDSDTFYNESFTNFVSKRADINSLNDESKRKKLEIKLKTÖL	1003
Db	1001	STTEMEKFEYENILKNNDTYFNDDIKQEVASNSKVITGLF--EQÖKNALNDEIKKLDTÖL	1059
QY	1004	SFDLYNNKYKLKLELPKPKKTATVGRYKMQIKLTLKQÖLESKLSNPNKHYLÖNFVSFF	1066
Db	1060	SFDLYNNKYKLKLELPKPKKLEÖDOKMQIKLTLKQÖLESKLSNPNHNYLÖNFVSFF	1119
QY	1064	NKKKEALIAETENULENTKLLKHVGLVYVYGGESSPKLTSEESIOEDMYASILENPK	1123
Db	1120	NKKKEALIAETENULENTKLLKHVGLVYVYGGESSPKLTSEESIOEDMYANDEKFR	1179
QY	1124	VLKLESKLDONLNEKKKLSLYSSGLHLHLIAELKLVINKNYTGNSPENNTGNVNALE	1183
Db	1180	VLKSTIDKLDNLDHLGKRRKLSFSSGHLÖITELKEVINKNYTGNSPENNNKRVNEMALK	1239
QY	1184	SYKFFLEPBGDIATVASE-----SGSOTLEQÖSPKPKASHVGAESNTITT	1229
Db	1240	SYENFLPE--AKVTTIVVTPPOPDYTPSPBLSVARVSGSGSTREFOITPTSGSLTLELÖOVOO	1298
QY	1230	SÖNVDDDEVADVIYPIRGESEEDYDÖLGOVVTGEATVPYIÖNITISKEINEVEVYLP	1289
Db	1239	LÖNVDEEDDSIYVPIRGESEEDNEYDÖVYVGEALSYT--MDNITISGFENEIVYILKPL	1357
QY	1290	AGVYRSLKKÖLÖENNVMTEPVNVAVDKLINSREKNREKNRVALESDLIPYKÖLUTSSNVVADP	1349
Db	1358	AGVYRSLKKÖLEKNIPFENLNLMDILNSRLKRRKYPFLDVLESDMQFKHISSENEYIDS	1411
QY	1350	YKFLNKERDKFSSYIVYKDSIDTÖDINFANDVLYGYIKLISEKYSDDLSDISIKKIY-----	1404

Db	1418	FKLINSEKNTILKSYTYIKESVENDIKFAQESISYIEKVLAKYDDLESIKKYIKEKE	14777
QY	1405	-----NDKQSENEKYLPFLNNIETLYKTVNDKIDLVJHLEAKVLNITYE	14449
Db	1478	KFPSSPTTPSPAKTDEOKRESEKFLPFLNIETTELNNLVNKIDYLIINLKAINDCAYE	15373
QY	1450	KSNVEVKTLELNKYTIQDKLADFPKNNNVGJADJSTYIINNILLTKPLSGWFMELA	15099
Db	1538	KDEAHVKTITSLDKAIDDKIDLPKNNNDDAIKTLINDYTKKMDLKGKLSGLV-QNP	15566
QY	1510	KTVSNLIDGNLQGMELNISOHCVKROCPONSGCFPHLDERECCCLNLYKQEGDCYEN	15699
Db	1597	NTIISKLEEGFQDMNLNISGHCVKROCCPENSGCFPHLDERECCCLNLYKQEGDCYEN	16566
QY	1570	PNPPTCENENGGCDADACTEEDSGSNMKTITCCTCPDPSYPLFDGIFCSSNFIIGISFL	16299
Db	1657	PNPPTCENENGGCDADACTEEDSGSNMKTITCCTCPDPSYPLFDGIFCSSNFIIGISFL	17166
QY	1630	ILMLILYSFI 1639	
Db	1717	ILMLILYSFI 1726	

RESULT	5			
ID	MSPI_PLAFM	STANDARD:	PRT:	1701 AA.
AC	P08569;			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)			
DE	(PMMSA) (P190).			
GN	MSP-1.			
OS	Plasmodium falciparum (isolate mad20 / Papua New Guinea).			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; plasmodium.			
OX	NCBI_TaxID=70153;			
	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88011243; Pubmed=3079521;			
RA	Tanabe K., Mackay M., Goman M., Scaife J.G.;			
RT	"Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum.";			
RL	J. Mol. Biol. 195:273-287(1987).			
	[2]			
RN	REVISIONS TO 1403; 1569 AND 1629.			
RP	Tanabe K.;			
RL	Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 1-115 FROM N.A.			
RX	MEDLINE=86136024; Pubmed=3004972;			
RA	Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,			
RT	Stunnenberg H., Bujard H.;			
RL	"Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merozoites: studies at the genetic level.";			
	EMBO J. 4:3823-3829(1985).			
	-1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (POTENTIAL).			
	-1- PVA: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42 KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation-			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC				
DR	EMBL; X05624; CA29112.1; -			
DR	PIR; A26868; A26868			
DR	PIR; B25120; B25120.			

DR InterPro: IPR000561; -
DR Pfam: PF00008; EGF 1.
KM Melaria; Mezozyote; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 607 607 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 919 919 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF6BA98 CRC64;

Query Match 60.23; Score 5069; DB 1; Length 1701;
Best Local Similarity 60.18; Pred. No. 6.3e-151;
Matches 1042; Conservative 226; Mismatches 339; Indels 126; Gaps 29;

QY 1 MKIIFFLCSFLFIINTOCYTHESYOELVKRLALDAVLGYSLFQEKWVNEGSGT 60
DB 1 MKIIFFLCSFLFIINTOCYTHESYOELVKRLALDAVLGYSLFQEKWVNEGSGT 60
QY 61 AVTTSTPGSGSVASGSGSVASGSGSVASGSGSVASGSGSGSRTPSDNSDS 120
DB 61 AVTTSTPGSGSVASGSGSVASGSGSVASGSGSGSRTPSDNSDS 114
QY 121 DAKSYADLKHVRNYLLTTELKLPOLFDLTNHLTLCDNHGKRYLDGEEINELLYK 180
DB 115 NTKYADLKHVRNYLLTTELKLPOLFDLTNHLTLCDNHGKRYLDGEEINELLYK 174
QY 181 LNFYFELLRAKLVANDVCANDYCOIPEFNKIRANEDLVKLVFGYRKLPLNDKDVGMED 240
DB 175 LNFYFELLRAKLVANDVCANDYCOIPEFNKIRANEDLVKLVFGYRKLPLNDKDVGMED 234
QY 241 YIKKNNKTININELIEESKKTIDKNNAKREBEKKLYOAOYDLSYKQLEBAHNLIS 300
DB 235 YIKKNNKTININELIEESKKTIDKNNAKREBEKKLYOAOYDLSYKQLEBAHNLIS 294
QY 301 VLEKRIDTLKKNKIRKLLDKINIK----NPPRANGNPNTLLD-KNKKIEHEKEIK 355
DB 295 VLEKRIDTLKKNKIRKLLDKINIK----NPPRANGNPNTLLD-KNKKIEHEKEIK 351
QY 356 EIAKTIFNIDSLTDPLELEYLREKNNKIDISAKVE--TKRESTEPN-EXPNGVTYPLS 412
DB 352 EIAKTIFNIDSLTDPLELEYLREKNNKIDISAKVE--TKRESTEPN-EXPNGVTYPLS 411
QY 413 YNDINNAL---NELNSFGDLINPFDYTKEPSKNITYDN-ERKKPINEIEKEIKLEKKI- 467
DB 412 LTIIDHNSLADNDKNSYDGLMNP-DTKEKINEKITDNKEKIKIINNKKQIDLEENIN 470
QY 468 ---ESDKKSYEDSKSLNDITTEYKELLNELYSKFNNDIDLNFEKMGKRSYKYEKL 524
DB 471 HTRKQNNKTLLEDEKES---KDYELLLEKFEYKMFNNNDKDVADLIFSRATYVNEKK 526
QY 525 THHTFASYSKSHNLEKTLKALKYMEDYSLRNIVVERKELKYNNLSKIENETLETVEN 584
DB 527 RYNNKESSSNNSYVNOYKLLKALSYLEDYSLRKGISKDRNHYTTLTGLEADKIKLTEE 586
QY 585 IKKDEQLFEKK---TKDENKDPDEKILEVSDIYKVOYVYLVNKKIDELAKTOLLKVN 641
DB 587 IKSEENKILKEKNFGKGLTHSANAS---LEVSDIYKVOYVYLVNKKIDELAKTOLLKVN 642
QY 642 ELKHNHVPNSYQENKQEPYLYLVKKEIDKLKVPKPYESLINEKKNKITEGSDNS 701
DB 643 QLADSIHVNPYRQNNPEPYLYLVKKEIDKLKVPKPYESLINEKKNKITEGSDNS 695

QY 702 EPSTGEITGQATTKPGQAGSALSGSVQAOEQKA-----QPPVPVPEAKAQ 754
DB 696 --ITQPLVASETTEDGGSTHTL-SQSGETEVEETEVEETGHTTYYITLPPKEES 752
QY 755 VPTPAPYNNKTEN-----VSKLDYLEKYLEFLMTSYCHKRYIIVSHSTANEKILKQ 806
DB 753 APKEKVVENSIEHKSNDNSQALKRYTYLKLDEFLRSYCHRYIIVSNSMQKLEEV 812
QY 807 YKTRKEESKLSGCDPLDLFNITONNIPWYSMFDLSLNSLSQLEPMETYEKWCNLYKL 866
DB 813 YNLTPEEKELKSCDPLDLFNITONNIPWYSMFDLSLNSLSQLEPMETYEKWCNLYKL 872
QY 867 KDNKIKLLEAKV-----STSVKTLSSS-----MQPLSLTPDQKEVS 908
DB 873 KEENHKKLEEQKQITGSTSTSSPGNTTYTAOSATHSNSQNOASNSNTNTNGAVS 932
QY 909 AND---DTSHS---TNLNSLKLENTLSGKKNNIYOEL-IGOKSENFYERILKDS 960
DB 933 SGPAVESHDPPLVLSISNDLKGIVSLNNGKNTKYPNPLISTTEMEKYEINILKND 992
QY 961 TFYNESFTNPKSRADDINSLNDESKRKKLEEDINKAKTILQSLFDLYNKYKLERLFD 1020
DB 993 TYFNDDIKQFVKSNSKYITGLT-ETORNALDEIKLKLDLQSLFDLYNKYKLERLFD 1051
QY 1021 KKTIVGKYKMOIKKLTLLKEQLESKLSLNNPKRYLVNFSVFNNKKKEAEIAETENTLEN 1080
DB 1052 KKKELGQKMOIKKLTLLKEQLESKLSLNNPKRYLVNFSVFNNKKKEAEIAETENTLEN 1111
QY 1081 TKILLKHYKGLVKNYNGESSPLKTLSESIQTEDNVASLENFKVLSKEGKLYKNDLNEK 1140
DB 1112 TKILLKHYKGLVKNYNGESSPLKTLSESIQTEDNVANLEKFRALSKIDGLNNDLHGK 1171
QY 1141 KKLISLSSGLHLLAELEKRYKNNKNTGNSPENNTYNNALSEYKFFLPCTGYAAYVS 1200
DB 1172 KKLISLSSGLHLLAELEKRYKNNKNTGNSPENNTYNNALSEYKFFLPCTGYAAYVS 1230
QY 1201 E-----SGSDTLEOSOPKKPASTHYGAESNTTTSQNDVEDVDYIVPIF 1246
DB 1231 PPGQDYTPSPSLAVSGSSSGSTKRETFQIPPTSGSLTELOQVOYQANDDEEDSLVPIF 1290
QY 1247 GESEDDYDGLQVVTGAVTPSVIDNLSKLENEYEVLTKPLAGVYRSLSKKOLENNVT 1306
DB 1291 GESEDDYDGLQVVTGAVTPSVIDNLSKLENEYEVLTKPLAGVYRSLSKKOLENNVT 1349
QY 1307 FNVAVKDIINSRFRKREKRYVLESDILPKYKDLTSSNIVYKAPKFLNKKERKDFLSSYN 1366
DB 1350 FNVAVKDIINSRFRKREKRYVLESDILPKYKDLTSSNIVYKAPKFLNKKERKDFLSSYN 1409
QY 1367 YIKSDIDINFDNFANDVGYKRIISEKYSKSDLSIKYI-----ND 1406
DB 1410 YIKSEVEDIKFAQEGSIYEKYLAKTKDDLESTIKYIKEKEKFPSSPPTPPSPAKTD 1469
QY 1407 KQGENEKYLPFLNIETLYKYTVNDKIDLEFVTHLEAKVLYNTYKESNVEYKIKELNYKTI 1466
DB 1470 EQKESKFLPLFLNIETLYKYTVNDKIDLEFVTHLEAKVLYNTYKESNVEYKIKELNYKTI 1529
QY 1467 QDKLADFRKNNNEVGIADLSTQYVHNNLLTKFLSTGAVFENLAKTVLSNLDGMLQMLN 1526
DB 1530 QDKLADFRKNNNEVGIADLSTQYVHNNLLTKFLSTGAVFENLAKTVLSNLDGMLQMLN 1588
QY 1527 ISOHQYKQKOPNGSGCFRHLDERECCCLNLYKQEBDKCVENNPICNENNGGCDADAK 1566
DB 1589 ISOHQYKQKOPNGSGCFRHLDERECCCLNLYKQEBDKCVENNPICNENNGGCDADAT 1648
QY 1587 CTEEDSGSGNKKTICETKTPDPSYPLFDGIFCASSNPLGISFLILMLILYSFI 1639
DB 1649 CTEEDSGSGNKKTICETKTPDPSYPLFDGIFCASSNPLGISFLILMLILYSFI 1701

RESULT 6
MSPL_PLAFF STANDARD; PRT; 1701 AA.

AC P13819;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
 DE (PMSA).
 GN MSP-1.
 OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID:5837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8142999; PubMed=2449612;
 RA Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,
 RA Brown G.V., Anders R.F., Kemp D.J.;
 RT "Variation in the precursor to the major merozoite surface antigens
 of Plasmodium falciparum";
 RL Mol. Biochem. Parasitol. 27:291-302(1988).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 (POTENTIAL).
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
 CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M19143; AAA29653.1;
 DR PIR: A54498; A54498.
 DR InterPro: IPR000561;
 DR Pfam: PF00008; EGF. 1.
 DR Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT CHAIN 1 19
 FT CARBOHYD 20 1701 POTENTIAL.
 FT CARBOHYD 110 110 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 607 607 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 919 919 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 1701 AA: 193719 MW: 3920875273D8552 CAC64;

Query Match 60.1%; Score 5065; DB 1; Length 1701;
 Best Local Similarity 60.1%; Pred. No. 8.3e-151;
 Matches 1042; Conservative 227; Mismatches 338; Indels 126; Gaps 29;

QY 1 MKTIIFLCSTFLPINTOCYTHESYQELVKKLEALDAVLGTGSLFQEKVNLNEGTSCT 60
 DB 1 MKTIIFLCSTFLPINTOCYTHESYQELVKKLEALDAVLGTGSLFQEKVNLNEGTSCT 60
 QY 61 AVTTPSGSGSVASGSGSVASGSGSVASGSGSVASGSGSGSRTPNSDSSD 120
 DB 61 AVTTPSGSGSVASGSGSVASGSGSVASGSGSVASGSGSGSRTPNSDSSD 120
 QY 121 DASSYADLKRVRNYLTITELKYPQLFDELTHMLTLCNDHGRKYILIDGEEINELLYK 180
 DB 115 NMTTYADLKRVRNYLTITELKYPQLFDELTHMLTLCNDHGRKYILIDGEEINELLYK 174
 QY 181 LNFYDLRAKLNDVCAVCOIPFNLIKIRANELDVLKLVFGYRKPLDNIKDNVGMED 240

DB 175 LNFYDLRAKLNDVCAVCOIPFNLIKIRANELDVLKLVFGYRKPLDNIKDNVGMED 234
 QY 241 YIKKNKTIENINELIEBSKTTIDKNNATKEEKKKLYQAQIDLSYNNKOLEAHNLIS 300
 DB 235 YIKKNKTTIANINELIEBSKTTIDKNNADNEEKKKLYQAQYMLFYNNKOLEAHNLIS 294
 QY 301 VLEKRIPTLKKNNENIKELLOKINEK---NPPANGSNPNTLLD-KNKKIEBHEKIK 355
 DB 295 VLEKRIPTLKKNNENIKELLOKINEK---NPPANGSNPNTLLD-KNKKIEBHEKIK 351
 QY 356 EIAKTIENIDSLPTDPLELEYLREKNKNIDISAKVE--TKESTEPN-EYPNGVTPYLS 412
 DB 352 EIAKTIENIDSLPTDPLELEYLREKNKNIDISAKVE--TKESTEPN-EYPNGVTPYLS 411
 QY 413 YNDINNAL---NELNSGDLINPDYTKREPSKNTYTDN-EKKKINIEKIKIEKKTI- 467
 DB 412 LTIHNSLAADNDKNSYDGLMNP-DTEKEKINEKLTTPNKKERKIFINNKKQIDLEEKIN 470
 QY 468 ---ESDKSYEDRSKSLNDITREYKELLNELTYDSKFNNNIDLTNFKMMGKRYSYVEKL 524
 DB 471 HTKQNKLLIEDYKS---KDIIELEKFEYKKNNNNDKDVYDIKPSARTYVNEKQ 526
 QY 525 THNTFASYSKNNLEKLTALKYMEDYSLRNIVREKELKYNNLSKTIENEIETLVEN 584
 DB 527 RYNNKFFSSNNSYVNOQKLLKALSYLEDYSLRGIISKDPNNHYTTLTGLEADIKLTEE 586
 QY 585 IKRNEBDFEKK---TKDEKDPKELLEVSDIYKQVQAVLNNKIDELKQIOLIKVY 641
 DB 587 IKSENNKILEKNFKGHSANAS---LEVSDIYKQVQAVLNNKIDELKQIOLIKVY 642
 QY 642 ELKNNIHVPNSYKOEKOEPEYLVTLKKEIDKLFVMPKVESLINEKKNITKEGSDNS 701
 DB 643 QLKDSIHVPNTIKQKNPEPYLVTLKKEVDKLEFIPKMDIKKQAVLSS----- 695
 QY 702 EPSTGEITQATTKPGQAQSALEGDSVOAQAQKQA-----OPVVPVPEAKAQ 754
 DB 696 --ITQPLVAASETTEDEGHSHTFL-SQSGEVEVEETEVEETVGHATTYVITLPPKESS 752
 QY 755 VPPRPVAVNNKTEK-----VSKLDYLEKLYEFLNTSYTCHKYLYIVSVSTNNKTLKQ 806
 DB 753 APKEVAVVENSIEKHSNDNSQALTKYVTLKKEIDKLFKSYCHKYLYIVSVSSMDOKLLEV 812
 QY 807 YKITEEESKLSGDPDLFLFNQNNIPVMYSMDLSLNSLSQLFMEIYKEMWCNLYKL 866
 DB 813 YNLPBEENELKSDPDLFLFNQNNIPVMYSMDLSLNSLSQLFMEIYKEMWYLYKL 872
 QY 867 KDNKIKNLLEBAKV-----STYKVLSSSS-----MQLSLTPQDKPEVS 908
 DB 873 KEENHIKLLLEBOKLITGSTSPGNTVTYNTAQSATHNSNQOQSVASSTNTONGVAVS 932
 QY 909 AND---DTSHS---TNLNSLSKLFENILSLGNKNKNYQEL-IGOKSSEPFYEKILKSD 960
 DB 933 SGPAVVEESHDPLTVLSISDLKGIYSLNLGNKTKVNPDLTISTEMERFYEYILKNND 992
 QY 961 TFYNESFTNEVSKADINSINDSKRRKLEEDINKLTKTLOLSFDLYNNKYKLEKLPD 1020
 DB 993 TYENDIDKQGVKSKSYITGLT-ETQKNALNDEIKKIKTLOLSFDLYNNKYKLEKLPD 1051
 QY 1021 KKTIVGKYKMQIKKLVLLKQLESKLSLNNPKHVLONESVFNNKKKEAIAETENTLEN 1080
 DB 1052 KKEELGQDKMQIKKLVLLKQLESKLSLNNPKHVLONESVFNNKKKEAIAETENTLEN 1111
 QY 1081 TKLILKHYKGLVYVYNGESSPLTSLSESIQTDNVAASLENFVLSKLEBKLDONLLEK 1140
 DB 1112 TKLILKHYKGLVYVYNGESSPLTSLSESIQTDNVAASLENFVLSKLEBKLDONLLEK 1171
 QY 1141 KKLVSLSGGLHLLIAELKEVYIKKNKNTGNSPSENNDVNNALSYKFLPEEGDVAVVVS 1200
 DB 1172 KKLVSLSGGLHLLIAELKEVYIKKNKNTGNSPSENNDVNNALSYKFLPEEGDVAVVVS 1230
 QY 1201 E-----SGSDTLEQSQPKKPASTHVGASNTITTSQNVDEVDVIVTIPF 1246


```

Db 639 SHHPNITKPKQKDEPYLYLVKKEVDKLEFIPKVKMLKKEQAVLSS-----IT 689
Qy 706 EGETTGQATKPGQAGSALGSDVQAOAOQOK-----AOPPVYPPVPEAKQOV 755
Db 690 QPVAASEETEDGHSHPTLSQSGEETEVEETETETGHTTVTTLTPKPKAVVENSJTH 749
Qy 756 PTPAPAPNKNKTKVSKLDYLEKLEFLNTSYCHKYLIVSHSTNNKELKQYKTKREES 815
Db 750 KS-----NDNSQALTKYVLYLKLDEFLTKYCHKYLIVSSNDQKILEYNTL-PEEN 803
Qy 816 KLSGCDPLDLFNQNNIPVWYSMEFSLNLSQLEMEIYKEMVCLYKLDKDKLNL 875
Db 804 ELKSCDRDLFLFNQNNIPVWYSMEFSLNLSQLEMEIYKEMVCLYKLDKDKLNL 863
Qy 876 LEEAKV-----STAKTSSSS-----MPLSTFQDKPEVAND---DTS 914
Db 864 LEEPKQITGTSSTSGPGTWTYTAQSAHSNSQNOQSNASTNTQNGVAVSSGPAVEES 923
Qy 915 HS-----TNLNSLKLFEIILSGKNNIYQEL-IGKSSSEYKELKDSPTFNESEFTN 969
Db 924 HDPITVLSISNDLKGIVSLNKGKTVKPNPLITSTEMEFYENILKIMPIFNDDIKQ 983
Qy 970 FVSKADINSLNDESKRKLEEDINKLKTLOLSPDLNRYKLELFDKRTVCKYK 1029
Db 984 FVKSNSVYITGLT-ETQKNALNDEIKKLDTLOLSPDLNRYKLELFDKRTVCKYK 1042
Qy 1030 MOIKLLFLKLESLKNSLNPKNHVLONFVFNKKKEALIAEENLENTKILKLYK 1089
Db 1043 MOIKLLFLKLESLKNSLNPKNHVLONFVFNKKKEALIAEENLENTKILKLYK 1102
Qy 1090 GLVYKNGESSPLKLTSEESIQEDNVAASLENFVYLSKLEKLDKNDLLEKKLSYSSG 1149
Db 1103 GLVYKNGESSPLKLTSEESIQEDNVAANLEKFPVLSKIDKNDLNLGKKLSYSSG 1162
Qy 1150 LHLIAELKEVYKKNKTGNSPENNDVNNALSKYKFLPEGDVATVSE----- 1201
Db 1163 LHLITELKEVYKKNKTGNSPENNNKVNALSKYENLPE-AKVTVVVPPOPDVYPS 1221
Qy 1202 -----GSDPTEGOSOPKRPASTVAGAESNTITTSQNVDDVDVYIYPIGESSEPD 1255
Db 1222 PLASRVGSSGSSSTKEEQIPIPSGSLTELOQVQLOLNDDEEDSVLPIGESSEPD 1281
Qy 1256 LGQVVTGEAVTPSVIDNLSKLENEEVLVLPKLAGVYRSKLOLNNVMTFNVNVDIL 1315
Db 1282 LDQVVTGEAVT-NDNLSGFEENEVDIYVLPKLAGVYRSKLOLNNVMTFNVNVDIL 1340
Qy 1316 NSRPNKRKRNKLVLESDILIPKDLTSSNYVVKDPYKFLNKKRKFSLSYITKDSIDT 1375
Db 1341 NSRLKRRKRYPLDVLSDLMOPKHAISNEXIIEDSFKLNSQKNTLLKSYKIKESVEND 1400
Qy 1376 INFANDVLYGYKILSEKSPDLSDISKYI-----NDQGENEKELP 1416
Db 1401 IKFQEGISYIEKVLAKYKDDLESTIKVYKEKEPESSPPTPPSPAKTDQKESKFLP 1460
Qy 1417 FLANNIETLYKTVNDKIDLEVLHLEAKVLYNTYKSNVEVKIKELNYLKTIDOKLADFEKN 1476
Db 1461 FLNIIETLYNNLVKIKIDYILNKAINDCVERDEAHVKITKLSDLKALIDOKIDLEFRNP 1520
Qy 1477 NNEVGIADLSTIDYVHNHNLKFLSTGMVFEVLAATVLSNLDGMLQGLMSIQHCYAKQ 1536
Db 1521 YDFEAIKRLINDTDKRMKGLSTGLV-QNFPNTIISKLITGEGFQDMLNLSQHCYAKQ 1579
Qy 1537 CPQNSGCFRHLDEEEKKCLINTYKQEGDKCYENPPTCNENNGGCDADAKTEEDSGSNG 1596
Db 1580 CPQNSGCFRHLDEEEKKCLINTYKQEGDKCYENPPTCNENNGGCDADAKTEEDSGSNG 1639
Qy 1597 KKINCECTKPDSPYLPFDGIFCCSSNPLGISPLILMLILYSFI 1639
Db 1640 KKINCECTKPDSPYLPFDGIFCCSSNPLGISPLILMLILYSFI 1682

```

RESULT 8

```

MSPL_PLAYO
ID MSPL_PLAYO STANDARD; PRT; 1772 AA.
AC P13828;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMSA) (230 KDA).
GN MSP-1.
OS Plasmodium berghei yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=5862;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90205979; PubMed=2320061;
RA Lewis A.P.;
RT "Sequence analysis upstream of the gene encoding the precursor to the
  major merozoite surface antigens of Plasmodium yoelii."
RL Mol. Biochem. Parasitol. 39:285-288(1990).
RN [2]
RP SEQUENCE OF 1093-1772 FROM N.A.
RX STRAIN-17XL;
RX MEDLINE=86124889; PubMed=2448778;
RA Burns J.M., Jr., Daly T.M., Vaidya A.B., Long C.A.;
RT "The 3' portion of the gene for a Plasmodium yoelii merozoite surface
  antigen encodes the epitope recognized by a protective monoclonal
  antibody."
RT Proc. Natl. Acad. Sci. U.S.A. 85:602-606(1988).
CC -1 SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
  (POTENTIAL).
CC -1 PPM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
  KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
  MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC CC
CC EMBL: J03612; AAA29762.1; -
CC EMBL: J04668; AAA29702.1; -
CC PIR: A28121; A28121.
CC PIR: A45532; A45532.
KW Malaria; Merozoite; Polypotein; Repeat; Signal; Glycoprotein;
  Transmembrane; GPI-anchor.
FT SIGNAL 1 18
FT CHAIN 19 1772
FT CARBOHYD 54 1772
FT CARBOHYD 54 406
FT CARBOHYD 646 406
FT CARBOHYD 829 406
FT CARBOHYD 1018 406
FT CARBOHYD 1090 406
FT CARBOHYD 1408 406
FT CARBOHYD 1446 406
FT CARBOHYD 1541 406
FT CARBOHYD 1629 406
FT CARBOHYD 1680 406
FT CONFLICT 1521 1521
FT SEQUENCE 1772 AA; 197230 MW; 9A6291658B0F45D CRC64;

```

Query Match 30.3%; Score 2548.5; DB 1; Length 1772;
 Best Local Similarity 32.4%; Pred. No. 1.3e-72;
 Matches 609; Conservative 331; Mismatches 589; Indels 349; Gaps 40;

```

Qy 1 KRIIFPLCSLFLPIINQCVTHESYQELVKKLEALDAVLTGSLFQKRNKVNLEGTSGT 60
Db 1 MKVIGLGFSEVFFAIKCKSETIEVYNDLIQKLEKLESISVDGELTFQKSOVIIN----- 54

```

```

QY 61 AVTTSTPGSKSVASGGSGSVASGGSVASGGSVSGSGNRRTPNSDSSDS 120
Db 55 -----ATOPFTIDPF 65
QY 121 DAKSVADLKHVRNVLTLIKELKYPQLFDLTNNHMLTCLDNIHGKRYLLIDGVEEINELLYK 180
Db 66 TNNHNA---QOVQDFVKFEGLGFTQETELVNLKALPNNYGVKYLIESKEEFGMLHA 122
QY 181 LNFYEDLLRANLVCANDYQCOIFENLKRANLVDLKLKFLGYRKPPLDNKDNVKKED 240
Db 123 INFYEDVLRDLNMCANNICEIPEHLKISEETEMLKVYLGRKPLENIQDDIEKLEI 182
QY 241 YIKRNKKTENINELIESKKT-----DKNKATKEEEKKLYQAOYDLSYKQ 291
Db 183 YIERNKETVALNMLIAEETKIQPEGNECDNDASCSDDKTKKKKPIYQAMNYIIFYKQ 242
QY 292 LEEHANLISYLERKIDPLKKNENKELLDKINEIKNPP-----PANSG 334
Db 243 LAETQKVEVLEKREVSFLKKNDKIRPLMOQIEVLNAAPVYTAETQIYTGSGSSTPGSGG 302
QY 335 N-----TPN-----TLDKN 344
Db 303 SSASGTSSSGASAGTQVEQANIVASVTYTPSGQNGEASTNPOTAOVYPVPTLLEKQ 362
QY 345 KRIEHEKEIKETIKFNIDSLFTDPLEEYLYREKNK---NIDISA--KVEETESTE 399
Db 363 KKIINGVLAQIKETIKFNIDSLFTDPLEEYLYREKNK---NIDISA--KVEETESTE 399
QY 400 P--NEYNGVYPLSYNDI-----NALNELNSTGDLINPDYKREPSKIYTDNERKKF 452
Db 423 PLTRYNGISYPLPNDYVYKRIANNAAE--TYGDLTHP--DNTPPLGDLATNQAERDL 479
QY 453 INEIKERIKETIKESIDSKSIEDRSKSLDITKEYEKLEINEIYDSKRNNDIDLTNEPKM 512
Db 480 IKAIKKIKETIKESIDSKSIEDRSKSLDITKEYEKLEINEIYDSKRNNDIDLTNEPKM 512
QY 513 MGKRYSVKELTNNHNTFASYENSKHNEKLTKAKYMEDSLNINIVEKELKAYKNLIS 572
Db 540 KTRDEYMTKK--TELNT--CEYGNFKELINKQNLQDLSLKKDIISNIEIEFNSKKK 597
QY 573 KIENEIETLVENIKDEBOLEFKKITQDENKPDKEILEVSDIYVQOVKVLNKKIDELK 632
Db 598 ELQYNIRLEAVQAKONVLA-----SKDVPPLSTLVELOIOLSKSLTQIOEOLN 646
QY 633 KTOIILKNVELKHNINHPNSKOENKOEPPYLLIYKKEIDOKLAKYPMKVESLIMEEKN 692
Db 647 KTEVSLKAOQLKDLVPKITGNKGPEPYLLIAVKKVEVDLAOFIPKIESMIKKERM 706
QY 693 K-----TEGQSDNSEPSTEGETGOATTKPGQAGSA-----LEG 727
Db 707 EOGPAITGESEEPVSGPSAESSTDRSSTSSSSSSSTPAALAESSATILPEAPAPAEA 766
QY 728 DSVQAOAOEOKOAPVPVPEAKAOPVPPAPVNNKTEVNSKLDYLEKLYEFLNYSYI 787
Db 767 ASPSTEAEEETTPPTQETOPQSOAASSTPAKPV-----MTKLYLEKLOKFLVFSYS 820
QY 788 CHKYILVSHSTNMELIKOYKITKEES--KLSCDPLDLFLFNIONNIPVYSMFDLNN 845
Db 821 CHKYVLQNSTINKDALSKALISEEKIRLAKCSLVDLALQNNMPVSLYESIVD 880
QY 846 SLQOLFMEIYEKEMVCLYKLD--NDKIKMLLEBA-----879
Db 881 GLQNIYVELYKEMVHIYKLDENPSIKSLVYKAGVIEPEVPAAPVPPVPAATEQOQOQ 940
QY 880 --KKVTSVATLSSSQOPLSLPQDKPEVSAANDTSHS-----916
Db 941 ATPDVQSDAPAPDSVQOQPEPTVSTTPEVTTSTPEASSAPGEGTPEGAGASCTEGATA 1000
QY 917 -----TNLNSLKLFENILS-----LCKNKIYELI 943
Db 1001 SNAATPAGTASGSAASNASTSVTPPAAAAAVPSTPAPAPAPPAANSOGNPDGICIR 1060
QY 944 GOKSSE-----NFEYKILKDSDTFYNESFTNFKSKADINDSLNDESKRKLKEE 992

```

```

Db 1061 SRAESEEDMPADDFELDNLKYSYLQOIDG--NNTFEFINFKSKKELIKALTPE--KVNOQLYL 1118
QY 993 DINKILKTLQSLPDLNKKYKLEKLEKLPKKTGVKKYKQKQIKLLILKEQLESKLSLNNP 1052
Db 1119 EIAHLKELSEHYDYRSYKYLKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEK 1178
QY 1053 KHAVLQNFVFNKKKKEADIAENTMENTKILKHYGLVYVYNGESSPLKLTSEESIQT 1112
Db 1179 FYILNGVNFNKKREAEKQYVDAKNTDMLKTYKARTKTYFTSEAVPLTSLKASIDR 1238
QY 1113 EDNYSLENFVLSLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEK 1172
Db 1239 ESNVLKIKETFRAYSRLERLRLKNNINLKERISYSGGLHHFEEFKELIKKODVTKKNP 1298
QY 1173 ENMTDVNNALESYKFLPEGTDAVTVVSESGDLEQSOPEKPASTHVAESNTI-----1227
Db 1299 DNAPEVTNAPEQYKELLPKGVATST--PAVAVTTTLADAPATPPEGAVPAGVAPGAVP 1357
QY 1228 -----TTSQNDDEVDVYIIVPIFESSEDDYDGLQVYTGAVPVSVIDNLSK 1276
Db 1358 GAVPAGVPGSGTDRVAGSSVDD-----NED--DDIYQIASGQSEDAPEKD--ILSE 1405
QY 1277 IENEYEVLYKPLAGVYRSLKQLENNVMPFNVVKILNSRFNKRFPKRVLESDLIPY 1336
Db 1406 FTNESLIVYTRKLSQYKSLKHHMLREKSTIKEDMTGNLNNKSKRNDPFLVLSHEDLF 1465
QY 1337 KDLTSSNVVADPYKFLNKKERKDFLSSYNIKDSIDTIDFANDVLYGYYKIIEKYSKD 1336
Db 1466 KDLSTNKVYINRNPQLDNDKDKQIYVLAATGKINMEDIETTTDGIKFNKAVELYNQ 1525
QY 1397 LDISIKRYI-----ND--KQENKRYLPFLNIETLVKTVMDKIDLEVIHLEAVLWYTY 1448
Db 1526 LAAYKEOATIEATNDPNNKKEKKYIPLEEDLGLVETVYGQAEESSEELQNLMDYKN 1585
QY 1449 EKSNVKIKELNLYKTIDOKLADF---KKNNFVGIADLSTYDNNHNLTKFLSTGMV 1504
Db 1586 EKAEFLITNKLLEYQIDEDLDFEVAENKKNHIAIA-----LNNLKNKGIV 1634
QY 1505 FENIAKTVLSNLDGNTQGM--LNIS--OHOCV--KKQCPONGCFRHLDERECKCLNLYK 1560
Db 1635 GEGESKRIAKML--NNDGMJLGLVDPKHCVDPRDIPKNAGCFRDNNGTEEMKCLGKY 1692
QY 1561 Q--EGDKVCENPNPCNNNGGCDADAKTEEDSGSNKKITCECTKXDSVPLFGICSS 1619
Db 1693 KGBGNTCVENNNPCTDINNCGCDPTASCQNAESTENSKITICTCKEPTPAVAYGVCCSS 1752
QY 1620 SNFLGISFLILMLILYS 1637
Db 1753 SSFNGSLILITLITLIVFN 1770

```

```

RESULT 9
MSPL_PLAFD
ID MSPL_PLAFD STANDARD; PRT: 233 AA.
AC P13827:
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMSA) (GLYCOPROTEIN 185) (GP185) (FRAGMENT).
OS Plasmodium falciparum (isolate CDC / Honduras).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
RX NCBI_TaxID=5836;
RN [1]
RP SEQUENCE FROM N.A.
RP MEDLINE=87106855; PubMed=3542719;
RA Howard R.F., Ardeshir F., Reese R.T.;
RT "conservation and antigenicity of N-terminal sequences of GP185 from
RL different Plasmodium falciparum isolates.";
Gene 46:197-205(1986).
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR

```

```

CC (POTENTIAL).
CC -1 PPM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M14632; AAB59227.1; -
CC PIR: A25814; A25814.
CC Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
CC Transmembrane; GPI-anchor.
CC SIGNAL 1 19
CC CHAIN 20 >233 MEROZOITE SURFACE PROTEIN 1.
CC CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC NON_TER 233 233
CC SEQUENCE 233 AA; 26050 MW; 1DE4D31ACAE965 CRC64;

Query Match 14.1%; Score 1187; DB 1; Length 233;
Best Local Similarity 95.1%; Pred. No. 2.8e-31;
Matches 232; Conservative 0; Mismatches 0; Indels 12; Gaps 1;

QY 1 MKIIFLCSEFFIINOCVTHESYOELVKKLEALDAVLTGYSLFQEKVNLNEGSGT 60
DB 1 MKIIFLCSEFFIINOCVTHESYOELVKKLEALDAVLTGYSLFQEKVNLNEGSGT 60
QY 61 AVTSTPGSKGSVASGSGSVASGSGSVASGSGSGNRPNPSNDS 120
DB 61 AVTSTPGSKGSVASGSGSVASGSGSVASGSGSGNRPNPSNDS 108
QY 121 DAKSYADLKHVRNYLTITKELKYPOLFDLTNHLTLCDNHFGEYLLIDGYEINELLK 180
DB 109 DAKSYADLKHVRNYLTITKELKYPOLFDLTNHLTLCDNHFGEYLLIDGYEINELLK 168
QY 181 LNFFFDLRAKLVNDVANDYCOIPFNKIRANEDLVKLVGTRKPLDNKDVGMED 240
DB 169 LNFFFDLRAKLVNDVANDYCOIPFNKIRANEDLVKLVGTRKPLDNKDVGMED 228
QY 241 YIKK 244
DB 229 YIKK 232

RESULT 10
MSP1_PLAFN STANDARD; PRT; 281 AA.
ID MSP1_PLAFN
AC P13820;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMMSA) (FRAGMENT).
GN MSP-1.
OS Plasmodium falciparum (isolate NF7 / Ghana).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=5842;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=88142999; PubMed=2449612;
RA Peterson M.G., Coppe R.L., McIntyre P., Langford C.J., Woodrow G.,
RA Brown G.V., Anders R.F., Kemp D.J.;
RT "Variation in the precursor to the major merozoite surface antigens
RT of Plasmodium falciparum."
RT Mol. Biochem. Parasitol. 27:291-302(1988).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -1- MISCELLANEOUS: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF

```

```

CC 83K, 42K & 19K ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M19144; AAA26354.1; -
CC PIR: B54498; B54498.
CC Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
CC Transmembrane; GPI-anchor.
CC SIGNAL 1 19
CC CHAIN 20 >281 MEROZOITE SURFACE PROTEIN 1.
CC CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC NON_TER 281 281
CC SEQUENCE 281 AA; 31226 MW; 6150B76E0F21B299 CRC64;

Query Match 13.0%; Score 1096.5; DB 1; Length 281;
Best Local Similarity 77.7%; Pred. No. 2.2e-28;
Matches 227; Conservative 8; Mismatches 32; Indels 25; Gaps 5;

QY 1 MKIIFLCSEFFIINOCVTHESYOELVKKLEALDAVLTGYSLFQEKVNLNEGSGT 60
DB 1 MKIIFLCSEFFIINOCVTHESYOELVKKLEALDAVLTGYSLFQEKVNLNEGSGT 56
QY 61 AVTSTPGSKGSVASGSGSVASGSGSVASGSGSGN-SRPTN----- 112
DB 57 EITR-----KG-ASAQSGASQSGASQSGASQSGSGSGSPSSRSRTLRSN 109
QY 113-----PSDSSDSDANSYADLKHVRNYLTITKELKYPOLFDLTNHLTLCDNHFGEY 166
DB 110 TSSGASPPADASDSDANSYADLKHVRNYLTITKELKYPESLDLPNHLTLCDNHFGEY 169
QY 167 LIDGYEINELLKLVNRYFOLLRAKLVNDVANDYCOIPFNKIRANEDLVKLVGTRK 226
DB 170 LIDGYEINELLKLVNRYFOLLRAKLVNDVANDYCOIPFNKIRANEDLVKLVGTRK 229
QY 227 PLDNKDVGMEDYIKKNTKININELIESKKTIDKNATKREKKKL 278
DB 230 PLDNKDVGMEDYIKKNTKININELIESKKTIDKNATKREKKKI 281

RESULT 11
YD86_SCHPO STANDARD; PRT; 1957 AA.
ID YD86_SCHPO
AC Q10411;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHEICAL 222.8 KDA PROTEIN C1F3.06C IN CHROMOSOME 1.
GN SPAC1F3.06C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxId=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA STEIN=972;
RA Connor R., Churche C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RA submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```


RA Kendrick K.E.; to the EMBL/GenBank/DBJ databases.
 RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-8 FROM N.A.
 RA Bai Y., Symington L.S.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI COMPLEX.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE ER AND THE GOLGI COMPLEX.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
 CC -1- SIMILARITY: BELONGS TO THE PIL5(TAP)/USO1/YBL047C FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb.sib.ch).
 CC -----
 CC EMBL: X54378; CA38253.1; -
 DR EMBL: L03188; AAB00143.1; -
 DR EMBL: U53668; AAB66559.1; -
 DR PIR: A38455; A38455.
 DR HSSP: P80220; IDIP.
 DR SGD: S0002216; USO1.
 KW Transport; Protein transport; Golgi stack; Cytoskeleton; coiled coil;
 KM Calcium-binding.
 FT DOMAIN 1 724 GLOBULAR HEAD.
 FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
 FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
 FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
 FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
 FT CA_BIND 475 486 POTENTIAL.
 FT CONFLICT 847 847 G -> E (IN REF. 2).
 FT CONFLICT 924 924 E -> K (IN REF. 2).
 FT CONFLICT 1253 1253 V -> I (IN REF. 2).
 FT CONFLICT 1319 1319 I -> V (IN REF. 2).
 FT CONFLICT 1461 1461 N -> S (IN REF. 2).
 FT CONFLICT 1581 1581 G -> S (IN REF. 2).
 FT CONFLICT 1600 1600 I -> V (IN REF. 2).
 FT CONFLICT 1661 1661 R -> S (IN REF. 2).
 FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
 FT CONFLICT 1772 1772
 SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

Query Match 5.8%; Score 486; DB 1; Length 1790;
 Best Local Similarity 20.6%; Pred. No. 1.5e-08;
 Matches 314; Conservative 277; Mismatches 562; Indels 368; Gaps 66;

QY 98 GSAVSGSGNSRTPNSDSSDSDAKSYADIKHRYNLYLTIKELKYPQLEFDLTNHHMTL 157
 DB 456 GDNAKENGSGNSKESKESDSDKDTGKDTGEYSGFKALFEVL-LNDVAELNLPKLEF 514
 QY 158 CDNHGKRYLDS--YEINL-----LYKNFEDLLRALN--DV--- 195
 DB 515 TTDLFFPFQDDHKYSEBLREITRNVTGNDLEDEEPLKAIQITSELTLTSTLAADIRP 574
 QY 196 -----CANDYCO-----IPNLKIRANLELV--LKKLVFGYGRKPLD 229
 DB 575 ISYTLFLIWMFGFKATNDPLSDKSVKLSLSTYQIQQDDVITIKCLVTMLLGAAYES 634
 QY 230 NIKDNVGMEDYIKNNKTT--ENINELIEESK-----TIKKKNA--TKEEK--KK 277
 DB 635 S-KSPSPRKEFYFEITIKTLGKNVYASRIKQFKKDSYFSDVMEDSLTPPELDETGLPK 693
 QY 278 LYQAOYDLSYKNOLEKHNLSVLEKRIQIDPLKKNENIKELLDINELKNNPPANSGNTP 337
 DB 694 YVFSTYFIQLFENI-----YRIATLASHDPEDEPISKIS----- 728

QY 338 NTLIDKNNKIEHEKEIKETIKFNIDSLFTDPLEVEYLREK-----KNIDISAKV 392
 DB 729 -----FEVEKLRQRCRK-LKGETTSIQETESTHEULTEKLALNHEKHEDEK 778
 QY 393 ETKESTEPNEYPCVYTPPLSYNDINNALNELNSFGDLINPDYKREPSKNITYDNKKF 452
 DB 779 QILNSSHSLSKENFSILETELKNVDSIDETQLRDLETKD--KENQTALL--EKST 833
 QY 453 INEKIKIEKKIE---SDKSYED-----RSKSLNDITKE---YKLNLEIDSKFNN 502
 DB 834 IHKQDSIKTLEKLETTLSQKKAEDGINKMGKDLPALESMQAVEBNCNLOKREKDS 893
 QY 503 NIDLTFEKKMGKRSYKVEKLTHTHTFASYENSKHNLEKTLKALKYMEDYSLNIVYK 562
 DB 894 NVNNOKEKSLKEDIKALKITIKAIN--ENLEBKICQNNLSKREH-----ISK 941
 QY 563 ELKYYK-----NLISKIENEIETLVENIKRDEQULEFKKITRDKPKDEKILEVSDIV 615
 DB 942 ELVEKSRFQSHDNLVAKLTETKLSLANNY-KDMQAEVESLIKAVESSKNESSIQLSN-- 998
 QY 616 KVQVQVILMKNKIDELKKTQILKNVELKHNHPNSKKQENQEPYLLVYKKEIDKL 675
 DB 999 -----LQNKIDMSQOE-----KENFQIERG-----SIEKNIEQLK 1028
 QY 676 VEMPKVESLINEEKNKIKTEGOSDNSEPTGEITGOATTKPGQAGSALGDSVQAOAO 735
 DB 1029 -----KITSDELQOTKEIIS--KSSDSKDEYSQI--SLKKEKETATANDENVKNISE 1079
 QY 736 EOKAQPPVPVPEAKAQVETPPAPVNNKTEVNSKLDYLEKLYEPLNTSYICHKYLVS 795
 DB 1080 LTKTRE-----ELBAEL--AAVKNLKNLETKLETSEKALKEVE----- 1117
 QY 796 HSTNEKTIKQKTKTPEESKLSGCDPLDLFNION-----NIPWMSMD----- 841
 DB 1118 -----NEEHLKEEKTQLEKEATEYEQOULNSLANLESKEHEHDLAOLKKEEIOANKER 1173
 QY 842 SLNLSOLFMEIYEKEMVNCNLYLTKNDKIKNLEBAKVKSTSVKTLSSSSMOPLSTP 901
 DB 1174 QYNELISQNLNEIISTQOENESIK-KKNDELEGEVKNMKSYSSEQSNKKSSEIDLALN- 1230
 QY 902 QDKPEVSANDTSTSTNL-----NNSKL-----PENLISGKKNNI 938
 DB 1231 QIK-ELKKNNTNEASLIESIKSYSESEYVKIKELQDECNKEFEVSELEDLKASDEKNS 1289
 QY 939 -YOELIOGKSENEYEKIKMDSTF-----FYNESFTNFKSADDDINSN--DESKR 988
 DB 1290 KYLEL--QKSEKIKIEEL--DAKTTELKIQLEKITNLSKAKESSELSRLKKTSSPERK 1345
 QY 989 KLEEDINKLKKRTQLDSFDLYNKYKLELRLFDKRRKTVGKYMQIKTLLEKQLESKXNS 1048
 DB 1346 NAEQOLEKIKNEIOLK-----NQAFEKER-----KLWBGSSJTIOEYSEKINT 1389
 QY 1049 LNNPKHVLQNFVFPNKKKEAEIETENTLENTKI---LKHKKGLVKYYNGESSPLK- 1103
 DB 1390 LEDELIRLQN-----ENELKAKEIINTRESELEKVSLSNDELLEEKQNTIKLSODEILSYKD 1445
 QY 1104 --TSEESIOR--EDNVAASLENFVLSKLBGKLDNLNLEKKKLSYSSGHHILIAELKE 1159
 DB 1446 KITRNDEKLISERQNRDRLDSLEKQLRAAOESKAKYEBGLKLEBESSEKAELEKSK 1505
 QY 1160 VIKNNKNTGNSPSENNTPDNNALNLSYKRFLEPTGDTAVTVVSESGSDTLEOSQOPKRPASTH 1219
 DB 1506 MMKRLKSTIES--NENELKSSMETIK-----SDEKLQES--KKSMEED 1545
 QY 1220 VGAESNTITTSQNVQDEVDYIIVPIGESEEDYDGLQVVTGAVPVSYIDNLSKIEEN 1279
 DB 1546 I-----KNLOHKSD--LISRIINESKEDIIEELKSKRIEAKGSELETVYKQELN 1593
 QY 1280 EYEVLYL-----KPLAGVYRSLK-KOLENNVMFNFNVVKDILNSRFNKNRPFKNVYL 1329
 DB 1594 AQEKIRINAENENTVLSKLEEDIERELDKQAE--IKSNQEKELLYSRLKELE----- 1644

```

QY 1330 ESDLIPIYKDLTSSVYVYKDPYKFLNKKERDKFLSSYNYIKDSIDTIDNFA-----NDVLGY 1385
DQ 1645 -----QELDSFGQAKQK-----SEERRAEVKKRFOYEKQSDLEKAMLLSTKYNDLVN- 1691
QY 1386 YKISEKYSKSLDSIKKTYINDKOGENEKYLEPFLNNIE---TLKYTVN-----DKIDLEFV 1436
DQ 1692 ---KEQAMKRDEDPVKKRTTDSQROEIEKLAKELDNLKAENSKLEKANESEIDMLMLV 1748
QY 1437 IHLKAVLNATYKESNVEVKI 1457
DQ 1749 TDDEKNAKTKSKLDLGEVI 1769

RESULT 13
RBP1_PLAVB
ID RBP1_PLAVB STANDARD; PRT; 2869 AA.
AC 000796;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
GN RBP1.
OS Plasmodium vivax (strain Belen).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
RT merozoites."
RL Cell 69:1213-1226(1992).
CC 1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC 1- SUBUNIT: HOMODIMER (POTENTIAL).
CC 1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M88097; AAA29743.1; -
DR HSSP; P36956; IAM9.
KW Malaria; Receptor; Signal; Transmembrane.
FT SIGNAL 1 17
FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
FT DOMAIN 18 2807 EXTRACELLULAR.
FT TRANSMEM 2808 2826 POTENTIAL.
FT TRANSMEM 2827 2869 CYTOPLASMIC.
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 2869 AA; 330213 MW; B9DBEA42205EBCFF CRC64;

```

```

Query Match 5.7%; Score 477; DB 1; Length 2869;
Best Local Similarity 20.9%; Pred. No. 4,7e-08;
Matches 372; Conservative 300; Mismatches 617; Indels 490; Gaps 87;

```

```

QY 254 --ELIEESKTTIDKNNKATKEEEKKLYQAQYDLSYKOLEBAHNILSVLEKRIIDLTK 311
DQ 1115 NYOYISEKVTILFKNNSTV-----YEAHSHINTVAHGI-TSNK 1133
QY 312 NE---NIKELDKIN-----EIKNPPANGNTPNTLLDRNK-IIEHEKEIKEI 357
DQ 1154 NEILSKYAEVEDKLNVLWQNDYKVKVKNENKQLEAIRGSMKSLKEVINHVSQMGLE 1213
QY 358 --AKTIKENIDSLFTDPLEEYLYREKKN----IDISAKY-----ETKESTEPNEIPNG 406
DQ 1214 STANTLAKSNAG-----KENHEDELELNKTGQMRDYEIKLKAIELEKGTV----- 1261
QY 407 VTYPLSYNDINNALNELNSFGDILNPDYTKPEKKNITTDNERKKFNLIEKIKIEKK 466
DQ 1262 -----NELKDAKEKNKVEP-----EPERNI-----IGHVLERITVEKDK 1296
QY 467 IESDKSYEDRSKSLNDITKYEKULNELIYDSKFNNNIDLTNPFKMMGRYSYK-VEKLT 525
DQ 1297 A--GKAYEE---MNSLKTIEKILQETSDS-QNELVTTSTIKHLENAGYEDVIKRN 1348
QY 526 HHNFPASYENSKHNLEKLTAKLTMEYSLR-----NIYVELEKYKULIS-KIENE 577
DQ 1349 EEDSIQLEKAK-SLETIDEMKKKLVOQVNMNIOGASIOGASIKLEMLKGVIELISTN 1407
QY 578 IETLYENIKKDE-----EOLFEKKITK--DENKPDEKILVSDIVKQVOKVLLMKNID 629
DQ 1408 YSILIEYKKNSSSVRSQLANGEFTKAECEKNASARLBAEKLEKQIYKDLIDYSDID 1467
QY 630 E-LKKTQULIKNV-ELKINI-----HVPNSYKQENKQEPYLL----- 664
DQ 1468 DKVKIKIEIKKELLKMKSSALTPEEESSEKFKOMCSHHEKN-KEGKKIIEYLNNDGGG 1525
QY 665 -----IVLKKEIDKLVE-----MKPYESLINE----- 687
DQ 1526 KANITDSQMEEVGVYVSKAEHAFHTVEAQVDTKAFCEIYAVYTKMDNLENSLEMKEVK 1585
QY 688 ---EKKNKITTEGSDNSRPSRTEGTGATTKPGQAGSALGSGVQAQAOEQAOQAPV 744
DQ 1586 VKCEKKNDEAKYKSKLKP-YDGR--KARVSENERKISEL-----KEKAVKKESSQL 1637
QY 745 PVVPEAKAQVTPPAVPANNKTEVNSKLDYLEKLYEFNTSYICHYLVASHSTNEKITL 804
DQ 1638 NDVSTRKSLQIDNCRQQLDSVLNIGRVK-QNMLQYRDSADKMKSVLPISELGAEKSL 1695
QY 805 KQYKITYEESKSLSCDPLDLFNIONNIPWYMSFDSLNSLSQLEMEIYEKEVNCMLY 864
DQ 1696 DKVKAKESEYK-----MLETVQNMEM---SRINVEEBSLTDIDKRTIDIEN--DLL 1741
QY 865 KTK---DNDKIKNLLEAKKAVTSVKTTLSSSQPLSLTPQDKREVSANDTJSHSTNLNN 921
DQ 1742 KMKROYEGGLQIKRENDKRSNFEVLGS-----EINALLDPSTSTPFIKL 1787
QY 922 SLKTFENILSLGKKNNI-YOELIGO-KSENFYEKILKSDTFYNESFTNFVKSKADDIN 979
DQ 1788 KLKEYDMGDL-KNYGVAMNEIHGEFTKSYNLITHLSNA-TDVSYVF-----EKAQSLR 1840
QY 980 SL---NDESKRKLLEEDI---NKLKK--TQLSFDLYNKKYKLEKLELDKKKTVGXYKMQ 1031
DQ 1841 ELAKKEEHLRRREEAFLFLNDIKKVESLTKLLEKMKKVAEYEGMKRDRHTVSQVLQVD 1900
QY 1032 IKKLTLLKEQLESKLNLSNNKHYLQNFVYFPNKKKAETAEIAE---TEMTLENTYILKHY 1088
DQ 1901 MKTIV---DELKT-LNDISECSSVLNNVSVYKVKESKADHYRRDANSYAESMTV--- 1952
QY 1089 KGLVKNYNGESSPLKTSEESITQEDNVAS---LENFVYLSKLEGLKLDNLTNLEKKKLSY 1145
DQ 1953 --LANYPLSDAKISSGMEFPAEKSNFKTDLELEITSVIS-----NSNELLKATIEQ 2002
QY 1146 LSSGLHMLIAELKVINKNKNTGNSPSENNTDVNNALSESRYKF--LPEGTDAVATVYSES 1202
DQ 2003 DSNVDIQKERESEQLAKDA-----TDIYVNVILKKNFMEKLEAKKKEEVYSEK 2051

```

```

QY 1203 GSDPL-----EOSQPKPASTHVGAESENTITTSQNVDEVD 1238
D 2052 VREALKRLSQVEGIRCHENFRLDNTNEELNKKMTITIRDKKSEESQLOKEMENMN 2111
QY 1239 DVT-----IVIFGESEDDYDL--GOVVTGEAVTPVIDNIIISKINEEVLV- 1285
D 2112 TYSNSTOLEGIVVASGESKEDIKLEKRSNEMERNISEKISTIDISKVTEMSTIDELVKL 2171
QY 1286 -----LKLPLAGYRSKL-----KOLENNVMPFNVAVKLDRFRKRENFVVL 1329
D 2172 GKNQOAHMISISTANNKTSKLLIMNKEKENTKCYD- IKDNSSSTGCVETLKGFTY 2230
QY 1330 ESDLIPKDLTSSNVYK--DPYKFLNKKRDKFSLSYNKTID-----SIDTDINF- 1378
D 2231 GSKL-----TFSSASEIYQNMADTYS-VNPAKHEK--ESLNALRDKLKELFLHONSISIV 2283
QY 1379 ---ANDVLGYKILISEKYSKSLDSIKKIYND-----KOGENEKYLIP 1416
D 2284 EGVQNMALALDKLNEE-KREMEDELYRNISETKLKOMEHSTDVFKPMELHGMNETNNK 2342
QY 1417 FLNNIEFLYKTVNDKIDLFVHLEAKVLYNTYK-----SNVEYKIKEL----- 1460
D 2343 SLKEKKKLVANDHMSMEKMKNGKLTYPESYQNNINITYIEAEVKLELDRDYG 2402
QY 1461 -NY-----LKTIDKLADEFKNNNFVGIADLSTDYNN----- 1491
D 2403 DNYGIVEBKQFSLIDRTNALMDITIEFKENNY-NLMEVNTETIRVNDYIEKTNK 2461
QY 1492 -----NNLTKFSLTGAVFEMLAR---TVISNLDDNLOGLN 1526
D 2462 LVQAKTEYQLENIKONDMLQNIPLKVSITIEFENVKRRKESILNDIYE--QERLLK 2519
QY 1527 ISQH-QCVKQCPONGSGFRHLDEREE--CKLNTYKOE 1562
D 2520 IGEHLDEKRVNVTETLSSY-EIDQKMEMSKNLEKSKS 2557

RESULT 14
MLP1_YEAST
ID MLP1_YEAST STANDARD: PRT; 1875 AA.
AC 002455:
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MYOSIN-LIKE PROTEIN MLP1.
GN MLP1 OR YKR095W OR YKR415.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=93247549; Pubmed=8483450;
RA Keelling R., Nguyen T., Chen E.Y., Botstein D.;
RA "A new yeast gene with a myosin-like heptad repeat structure.";
RT Mol. Gen. Genet. 237:359-369(1993).
RN [12]
RP SEQUENCE FROM N.A.
RX MEDLINE=94205265; Pubmed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RA "The complete sequence of a 15,820 bp segment of Saccharomyces
RA cerevisiae chromosome XI contains the UBI2 and MLP1 genes and three
RT new open reading frames.";
RT yeast 9.1349-1354(1993).
RL yeast 9.1349-1354(1993).
CC -FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC REPAIR.
CC -SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MLP1".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation

```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; L01992; AAA34783.1; -
 DR EMBL; X73541; CAA51948.1; -
 DR EMBL; Z28320; CAA82174.1; -
 DR PIR; S38173; S38173.
 DR SGD; S0001803; MLP1.
 KW Myosin; Heptad repeat pattern; Coiled coil; DNA repair.
 FT DOMAIN 69 487 COILED COIL (POTENTIAL).
 FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
 FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
 FT CONFLICT 301 301 R->A (IN REF. 1).
 SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C906867 CRC64;

Query Match 5.6%; Score 474.5; DB 1: Length 1875;
 . Best Local Similarity 20.7%; Pred. No. 3.5e-08;
 Matches 346; Conservative 289; Mismatches 545; Indels 489; Gaps 78;

```

QY 111 TNPDSNDS--DAKSYADL-----KRVNRYLLT-IKELKYPQLFDLTNNML 155
D 368 TDEADEDEHNSAKSSDFIFLKQLIKERRTEHLQOETIYELE-----HKV 418
QY 156 TLCDNNGFKRLIDGYE-EINELLYKLNFFYDILLRAKLVDCANDY---COIFNKLIR 210
D 419 PT---INSFKERTMLELENLNAALLLEHTSNERNKNAVYKELNANQKLECE----- 467
QY 211 ANELDLVLLKLVGYRKPDLNI-----KDNVGMEDYIKKKNKTIENINELIESKRTI 263
D 468 -NDLOTLLKQGLDLCROQYVLLITNSVNSDKPL-----RKKEIQFIQIMEDDSTI 520
QY 264 DKNNAKTEEEKKKLYQAQDLSTYKNQLEBAHNLISVLEKRRIDTLK-----KN 312
D 521 -----TESDSQK-----VTERLYEFKNIILOEKRNAELLVNRLADLESKE 564
QY 313 ENIKELDKI-NEIKNPPRANSNTPTLLDKNKKIEHEKEIKETIFNIDSLPTD 371
D 565 KKSQSLQKIESETVNEAKKAITLKSEKMDLESRIEQLQLEEL-RTSVPNEDASIS- 622
QY 372 PLELEYLRKKNKIDISAKVETKESTEPNEYPNGVTPPLSYNDINNALNLSFGDLIN 431
D 623 -----NVTIKQLTETKRDES-----QVQDLQRRISQI----- 650
QY 432 PFDYTKPESKNIYDNERKKFIDNEIKIKETKKIESDKKSYDRSKSLNDITKEYEKL 491
D 651 ---TRESTENMSLNLKEIDQLDLYSKSDISIKLKEKSSRIAEERKFL----- 696
QY 492 LNEIYDSKFNNNIDLTNPEK-NMGKRSY-----KYEKLTHNTPASYENKHNLEKL- 543
D 697 -----SNTDLITRAENDQLRKREDYLONTILKQDSKT-HETLNEYVSCSKSLIVE 746
QY 544 TKALKMEDYSLRNIVVEKELKYKKNLSKIENEITLV--ENIKRDEQLFPEKTKTKD 600
D 747 TELNLKEEQKLR-VHLEKNLKOELNKLSPKSDSLRIMVQLQTLQEREDLLETKSC 805
QY 601 ENKPEKILEVSDIYKVOVK-----VLANKIDELKK-----TQILK 639
D 806 QKKIDLEDALESLKETSQKDHKIKOLEEDNNSNIEMYOQKIKALKKDYESVITSYDSK 865
QY 640 NVELKHNHIVNSYKQENKQEPYLLV-----LKKELDKLVFAPKYESLINE 687
D 866 QTDIEKLQYKVSLEKLEEDKIRLHYVNMDETINDSLKELKEKSLINTDAYSQIKE 925
QY 688 EKKNIKTEGOSDNESEPTGEITGOATTKPQOAGSALGSDSYOAOQEOKQAPVPVP 747
D 926 YKDLFTTSSQ-----LQQTNSKLD-ESFDFDNLQIKNL----- 958
QY 748 VPEAKQVPPPPAVNNKTENVSKLDYL-EKLYEFLNTSYTCHRYLIVSHSTNKEKL-- 804

```



```

Db 959 -----TDEKTSLEDKISLKEQMFNLNDELQKKMEKAKFKKRISIL 1004
Oy 805 ----KOYKITEE-ESKLS-----SCDPLDLFN-IONNIPMYGMSFDSLNLSQLEFMEI 854
Db 1005 ONNKEVANKSEYESKSLQNDLDOOTIYANRANNYEOELOKHADSVKTSISELRQOL 1064
Oy 855 YEKEMVONLYKIKNDXIKNLLEAKKVSYSV-----TLSSSMQPLS-----LTPQ 902
Db 1065 HTYGQVKTNL-SRDQLEMLKENEKSMWSQKESLLEQDLDSNRIDELSQKKLTYDQ 1123
Oy 903 DKPEVSANDTSHTN--LNNSLKLEFNILSLCKNNIY--QELIGKSSSENFYEKI- 955
Db 1124 IQITLADKVENNSTNGPLNNIL-----ITLRREDILDTKYTVARDAKMLRQKISL 1177
Oy 956 ----LKSDPFYNSFT-----NFVSKRADDINL-----NDESKR 987
Db 1178 MDVELQDPRKLDMSRVEKENHSSIIQOHDIMEKLNOLNLSRNSITLRLNELENNKK 1237
Oy 988 KLEBEDINKLKTLL-----QLSFDLYNK-YKLELRLDKKKTKVGYKMOIKLITLL 1038
Db 1238 KELOSELDKLKONVAPLESSELTALKYSMOEKEQELK-----AAEYHRRKKRSODILEK 1292
Oy 1039 KEQLES--KLNS-LNNPKHVLQN-----FSVFENK-KKEAE-----TAE 1073
Db 1293 HEQLSSDYERLESEIELEKLEKNEROGAEBEKFNRRLRQOERLTKSLQSDSLTE 1352
Oy 1074 TENTLEMTKILKTKYKGLVYKNGESSPLKLSPESTIOTEDNTYSLNFKYL--SKLEG 1130
Db 1353 QVNSLRKAKVLEN-----SLSEANARIEE--LONAKAAQGNQMLEA 1392
Oy 1131 --KIKDN-----LNLEKKRLSYLS--SGLHLIAELKEVIR--NKNYTGNSPS 1172
Db 1393 IRKIQEABAKASRLQAKLEBSTITSTYESTINGLNEEITTLKEELEKORQIOQLOATSAN 1452
Oy 1173 ENNTDVNNALESYKPLPEGTIVATVVSSEGSPTLE--QSOPRKPASTHGAESNTITT 1229
Db 1453 EQN-DLSNIVESMKKSFEE--DKIKFIKEKTOEVNEKILEAQERLNDPSNINMEIKKKW 1509
Oy 1230 SQNVDEVDVYIPIFGESEDEYDGLQVYTGAVTPSVYDNLSTIKENEVLYYKPL 1289
Db 1510 ESEHEQVESOKI--REAEELAKKRIRLPTKEKIN-KILERKKEELEKEFEKKEVERE 1563
Oy 1290 AGYRS-----LKKOLENNVMTFENVVKDILNSRPNKRENFKNVLESDDLTPYDLTSSN 1343
Db 1564 KSMQOSEIDVYLRKQLEAKVQ--EKQKLENEYNK-----LOBEL--KDVPHSS 1610
Oy 1344 YVVKDPYKFLNKKERKDFLSSYNIKDSITDINDFANDVLGYKYLSEKYSKSDLSIKKY 1403
Db 1611 HISDD-----ERDK-----LRAEIESR-----LREFEANNELQAIKKK 1642
Oy 1404 INKQGENEKLPLPLNNIETLYKTVNDKIDLFVHLEAKVLYNTEKSNVEVKTIKELNYL 1463
Db 1643 SFD-EGKO-----QAMMKTT-----LLERK-----L 1662
Oy 1464 KTIDKLADEFKKNNNPFIADLSTDYNNHNNLTKFLSGVWFENLAKTYVLSNLDGNSLG 1523
Db 1663 AKMSQSLSEKQSE-----SPPKSVNNOVONPLGLPKRIEENSNSPPPLLSG----- 1711
Oy 1524 MLNLSHQCVKQCPQNSGCFRLHDEREECKCLNLYKQEGDKCVENPNP 1572
Db 1712 -----EKLTKLNKSSSGGNPFTSPSPKHLQNDNDKRSLSANKTDP 1754

```

```

RESULT 15
RBP2_PLAYB
ID RBP2_PLAYB STANDARD; PRT; 1251 AA.
AC 000799;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE RETICULOCYTE BINDING PROTEIN 2 (FRAGMENT).
GN RBP2.
OS Plasmodium vivax (strain Belem).

```

```

CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
  merozoites".
RL Cell 69:1213-1226(1992).
CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND (PROBABLE).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M88098; AAA29744.1; -.
KW Malaria; Receptor; Membrane.
FT NON_TER 1
FT NON_TER 1251
SQ SEQUENCE 1251 AA; 143741 MW; 54BA51C740AC572 CRC64;

```

```

Query Match 5.4%; Score 457.5; DB 1; Length 1251;
Best Local Similarity 21.1%; Pred. No. 7,7e-08;
Matches 306; Conservative 228; Mismatches 458; Indels 455; Gaps 71;

```

```

Oy 272 EEEKKKLYQAOYDYSYKQLEAHNLISYLEKRIDTLKKNENIKELDKITENIKNPPRA 331
Db 1 EFDKEKVDTSFDEK--KKSIEKAY-----EKKMGNTLKE-----LEKMDKEN-- 41
Oy 332 NSGNTPNLIDKNNKIEHEKEIKETIKTFNID--SLTDPLLEYLYRER--NK 384
Db 42 -----TEKEVEEAQIOYKRIF-----IDHDVNLNDEVEKSKIMEKITELYKK 84
Oy 385 NIDISAKVETKESTEPNYPNGVT-----YPLSYNDINNALNELNSFGDLINPPTYKEPS 440
Db 85 EID-----EIQKTI--NRYKQGDTSNFPYTYEOYNSATQSKAKIEDFINAIATTKGTSPTS 137
Oy 441 KNITDNERKKFNIKEIKIEKIKIEKKIESDKSYEDRSK--SLNDITKEYEKLIN--EI 495
Db 138 QDI--NE--LESIKEVEHKNLOLVKQESNSMEEMKQILSMKDLL--ILNNSET 185
Oy 496 YDSFNNNIDILTFNEKMMGKRYKVEKLTHTNFPASVENSKNHLEKTLAKTYMEDSL 555
Db 186 IAKETISNNT-----QNALGFEENKTKLNTKIDELLO----- 216
Oy 556 RNIVVEKELKYKKNLISKIENEIEFTLVENIKKDEEQLFEKKITGDENKPRDEKILEVSDIV 615
Db 217 RVAMMIEBAKAKH-----NNIDIALEDAQID-----TEVSKIE 249
Oy 616 KVQVQKVLNKKIDELKKTQILKKNVELKHNHVPNSYKQENKQEPYVYLIVLAKKEIDIK 675
Db 250 QINRE--INNKKDEIKSYLSIEIKYKDKCTEITSNKRQDK--IEPLE 294
Oy 676 VFMKRVES-----LINEKKNIKTEGQ-----SDNSPSTEGEITGOATTKPQOQAGSA 724
Db 295 KFKPNEESNSKVNINENINENIRNSDYLDKIDEAERKQASTKVELFHFHEH--TISNIF 351
Oy 725 LEGDSVOAQAOEQKQAPVPVPEAKAQAQVTPPAPVNNKTENVSKLDYLEKLYEFLNT 784
Db 352 KESRILGVEYTSQKKI-----NKAEDIMK--ELER----- 379
Oy 785 SYICHKTYLVSHSTMENKILKQYKTKREESKLSGCDPLDLFNIIQNNIPYV--SMFDSL 843
Db 380 -----HNSEIQTQVKGFOENLKNL-----NEPHANYNAEDL 411
Oy 844 NNSLSQLEFMEIYEKEMCNLYKLDKN-DKIKNLLEAKKAVTSVKTLSSSMQPLSLTPQ 902

```

Db 412 NNDKSTNAKVLE-----TNLESVKNHLSLITNIOGGEIKYSKAKDI-----MOKIKATSE 463
QY 903 DKPEVS---ANDDTSHTNINNSLAKFENILSLGKN-----NIYOLLIGOKSSEN 950
Db 464 NTAKEITLKVADQDSNVYVNLQITTEBNLIVTEKNRLNGIDSTITNIEGAL---KESKG 520
QY 951 FYE-----KILKSDTFYNESTFNVKSKADDINSLAND-ESKR 987
Db 521 NYELGFLEKLEELIGKNRKLKVDITKKSINSTVGNFSSLFNNFDLNOYDFKNINDYENKM 580
QY 988 ---KLEBEDINKLKTLOLSFDL---YNYK-KLE-----RIPDKKTGVKYYKMOIK 1033
Db 581 GEIYNEFEGLNKISENIRNASENTSDNSAKTLREAOKEKENVLNKEEFANKYLRDVK 640
QY 1034 KLTL-----LKEOLESLNLSL-----NNPKHVLONFVSFFNKKKEAEIAE-- 1073
Db 641 KVESFRFLFNKESLD-KINEMIKKEOLTVEGHGNVQOLVENIKELVDENNLSIDLKQA 699
QY 1074 -----TENTLEN-TKILKHVGLVYVNGESSPLKTLSEESIOTEDNYASLENF 1122
Db 700 TGNKEIOLKITHSLKNKAKILGHVDSAKYVGKITPELALTE----- 744
QY 1123 KVLKLBGLKDNINLEKKKLSYSSGLHLIAELKEVYKKNKNTGNSPSENNTDVN-NA 1181
Db 745 -LIGDAKLKTPAOELKPFESK-----NNVLETEMNSKN-----TNELDVHKNI 785
QY 1182 LESYK---KFLPEGDVATVVSSESG-----SDTLEQSQPKPA-ST 1218
Db 786 QDAKVALEILIAHDEIDTKOKDSSKLIEMGNQIYLVKVLINQIKNKISSIKSKEAVSV 845
QY 1219 HVG-----AESNTITTSQNVDEVDVYIPIFGESEEDYDLGOVVTGEAVTPSYIDN 1272
Db 846 KIGVNSKKHSELKITES---DKSYDNIAL---EKQTELOLNRSPTQEK-TINTSDS 897
QY 1273 ILSKIENEYEVL-YLKLPLAGVRSIKKOLEN--NMTFNVNVKDILNSRPNKRENEKNV 1328
Db 898 KLERIKTDVESLKNALKTLEGEVNALKASSDNHSHVQSKSEVPNPAL-SEIEKEET--- 952
QY 1329 LESDLIPYKDLFSNYYVVKDPYKFLNKKRDKPLSSVYIKDSI-----DTDI--NEAN 1380
Db 953 -----DIDSLNTALDE---LKKGRICEVSRKLIKDVYKTEISDDTELINTIEK 999
QY 1381 DVLGYVYKILSEYK---SDLDISIKYINDKOGENEKYLPE-----LNNIETLYKTV 1428
Db 1000 NVKAYLAYIKKNYEDQVDVLTLENEHFTKQVSHNEPTFNKSSSEELTKAVTDSKTI 1059
QY 1429 NDKIDLVIHLEAVLVNTYEKSNVEYKIKEL---NYLKTIOKDLADFKKNNNFVGIAID 1484
Db 1060 ISKLKGVYIE---VVENETEMNTIESSAKIEIALYNELKKNKTSLMEIYOTSNEVKLOE 1114
QY 1485 LSTQYNNHNLTKFLSTGWFENLAKTVLSNL-----LDGNLOGMLNISQH- 1530
Db 1115 MKSNAD-----KIYDVSKIFNTVLOTOKSNITVNOHSINNWKDLKGLKLOELTDADSSF 1168
QY 1531 --QCVRK 1535
Db 1169 TLESIRK 1175

Search completed: October 27, 2001, 15:57:18
Job time: 248 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2001, 15:49:05 ; Search time 36.52 Seconds
(without alignments) 3418.677 Million cell updates/sec

Title: US-09-269-874-3

Perfect score: 8424

Sequence: 1 MKIIFELCSFLFEIINTQCV.....SNLGLISFLILMLILYSFI 1639

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
1: PIR-68:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	8424	100.0	1639	2	S05603	major merozoite su
2	8191.5	97.2	1640	2	A24594	probable major sur
3	7965	94.6	1631	1	SAZOK1	major merozoite su
4	5136.5	61.0	1726	1	SAZOGM	major merozoite su
5	5121.5	60.8	1726	2	A45948	major merozoite su
6	5065	60.1	1701	2	A54498	major merozoite su
7	5053	60.0	1701	2	A26868	major merozoite su
8	3093.5	36.7	1751	2	A45604	major blood-stage
9	2999.5	35.6	1086	2	S16752	major merozoite su
10	2959	35.1	1726	2	A39401	merozoite surface
11	2829.5	33.6	1060	2	S06286	major merozoite su
12	2548.5	30.3	1772	2	A45532	major merozoite su
13	2509	29.8	1785	2	A45346	major merozoite su
14	2146	25.5	651	2	S47282	merozoite surface
15	2101	24.9	400	2	A45545	major merozoite su
16	1187	14.1	233	2	A25814	glycoprotein 185 -
17	1135	13.5	680	2	A28121	major merozoite su
18	1096.5	13.0	281	2	B54498	major merozoite su
19	1021.5	12.1	636	2	A45949	merozoite surface
20	815	9.7	168	2	A25522	major merozoite su
21	710.5	8.4	146	2	S03290	glycoprotein, 190K
22	659.5	7.8	138	2	P00125	major merozoite su
23	606.5	7.2	1939	2	T18372	repeat organellar
24	605	7.2	2401	2	T28676	riophy protein -
25	568.5	6.7	2266	2	T28677	riophy protein -
26	559.5	6.6	1979	2	C71622	hypothetical prote
27	529	6.3	115	2	B25120	major merozoite su
28	521.5	6.2	3394	2	T18501	hypothetical prote
29	507	6.0	2166	2	G70163	hypothetical prote

30	501.5	6.0	1957	2	T38077	hypothetical colle
31	495.5	5.9	284	2	A45518-	merozoite surface
32	487	5.8	1790	2	S67593	transport protein
33	483	5.7	1127	2	T28317	ORF MSV156 hypote
34	477	5.7	2829	2	A42771	reticulocyte-bind
35	477	5.7	3724	2	T18427	hypothetical prote
36	474.5	5.6	1875	2	S38173	myosin-like protei
37	465	5.5	476	2	B44997	merozoite surface
38	460	5.5	482	2	A44997	merozoite surface
39	457.5	5.4	1252	2	B42771	reticulocyte-bind
40	453.5	5.4	2485	1	H71621	serine/threonine-s
41	448	5.3	1650	2	T18444	hypothetical prote
42	443.5	5.3	2663	1	S28261	centromere protein
43	442.5	5.3	3973	2	B71612	hypothetical prote
44	440	5.2	1805	1	A64224	hypothetical prote
45	438	5.2	1738	2	T14867	interactin - slime

ALIGNMENTS

RESULT 1

S05603 major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
N:Alternate names: gp195 surface antigen
C:Species: Plasmodium falciparum
C>Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jun-2000
C:Accession: S05603; S04850
R:Myler, P.J.
submitted to the EMBL Data Library, April 1989
A:Reference number: S05603
A:Accession: S05603
A:Molecule type: mRNA
A:Residues: 1-1639 <MYL>
A:Cross-references: EMBL:X15063; NID:99896; PIDN:CA93163.1; PID:99897
R:Myler, P.J.
Nucleic Acids Res. 17, 5401, 1989
A:Title: Nucleotide and deduced amino acid sequence of the gp195 (MGA-1) gene from P1
A:Reference number: S04850; MUID:89345116
A:Accession: S04850
A:Molecule type: mRNA
A:Residues: 1504-1639 <MYL2>
A:Cross-references: EMBL:X1063
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1639/Product: major merozoite surface antigen #status predicted <MAT>

Query Match	100.0%	Score 8424;	DB 2;	Length 1639;
Best Local Similarity	100.0%	Pred. No. 6.2e-254;		
Matches 1639;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
OY	1	MKIIFELCSFLFEIINTQCVTHESYOELVKKLEAEDAVLNGYSLFQEKWLNCGTSGT	60	
DB	1	MKIIFELCSFLFEIINTQCVTHESYOELVKKLEAEDAVLNGYSLFQEKWLNCGTSGT	60	
OY	61	AVTSTPGSKGSVASGSGSVASGSGSVASGSGSVASGSGSVASGSGSRRTPNSDSSDS	120	
DB	61	AVTSTPGSKGSVASGSGSVASGSGSVASGSGSVASGSGSVASGSGSRRTPNSDSSDS	120	
OY	121	DAKSTADLKHRVRYLTLTKELKYPQLDPLDNNHMLTLCDDNHGFRYLLDGYEELNELLYK	180	
DB	121	DAKSTADLKHRVRYLTLTKELKYPQLDPLDNNHMLTLCDDNHGFRYLLDGYEELNELLYK	180	
OY	181	LNFFYDLRAKLNIDVCANDYCOIPNLEKIRANELDVLKLVFGYRKPDPNDIKDNGKMD	240	
DB	181	LNFFYDLRAKLNIDVCANDYCOIPNLEKIRANELDVLKLVFGYRKPDPNDIKDNGKMD	240	
OY	241	YIKKKRTIENINELIEESKRTIDKNKNAKFEKKKKLYQAOYDLSTYNNKOLEAHNLIS	300	
DB	241	YIKKKRTIENINELIEESKRTIDKNKNAKFEKKKKLYQAOYDLSTYNNKOLEAHNLIS	300	

Qy 301 VLEKRIIDLTKNENIKELLDKINEIKNPPANSNGTPTLLDKNKIEHEKEIKETAKT 360
Db 301 VLEKRIIDLTKNENIKELLDKINEIKNPPANSNGTPTLLDKNKIEHEKEIKETAKT 360
Qy 361 IKFNIDS.LFTDPLELEYLREKKNKIDISAKVETKESTEPENYENGVTYPLSYNDINNAL 420
Db 361 IKFNIDS.LFTDPLELEYLREKKNKIDISAKVETKESTEPENYENGVTYPLSYNDINNAL 420
Qy 421 NELNSFGDLINPFDYTKPSKNITTDNERKKFINEIKKIKIEKKKIESDKSYEDRSKS 480
Db 421 NELNSFGDLINPFDYTKPSKNITTDNERKKFINEIKKIKIEKKKIESDKSYEDRSKS 480
Qy 481 LNDITKEYEKLINLEYDSKFNNNIDLTNEKMKGRYSYKEKTLHNHTFASYSKSNL 540
Db 481 LNDITKEYEKLINLEYDSKFNNNIDLTNEKMKGRYSYKEKTLHNHTFASYSKSNL 540
Qy 541 EKLTKALKYMEDYSLRNIVYERKELKYKNLSKISKENIETLVENIKKDEQLFEKKIKTD 600
Db 541 EKLTKALKYMEDYSLRNIVYERKELKYKNLSKISKENIETLVENIKKDEQLFEKKIKTD 600
Qy 601 ENKPEDEKILEVSDIYKVOVOKVLLMKNIDELKKTQJLKNVLEKHNHVPNSYKOEKOE 660
Db 601 ENKPEDEKILEVSDIYKVOVOKVLLMKNIDELKKTQJLKNVLEKHNHVPNSYKOEKOE 660
Qy 661 PYYLIVLKEIDKLVKMPKVESLINEKKNITKEGSDNSEPTEGEBITGOATTKPGQ 720
Db 661 PYYLIVLKEIDKLVKMPKVESLINEKKNITKEGSDNSEPTEGEBITGOATTKPGQ 720
Qy 721 AGSALGEDSVQAOEQOQOAPVPVPYPEAKAOYPTTPAVNNKTENVSKLDIYEKLYE 780
Db 721 AGSALGEDSVQAOEQOQOAPVPVPYPEAKAOYPTTPAVNNKTENVSKLDIYEKLYE 780
Qy 781 FLNLSYIOCHKIYLVSHTSMNEKILKQYKITKEESKLSGCPDLDFINONNIPMYLMF 840
Db 781 FLNLSYIOCHKIYLVSHTSMNEKILKQYKITKEESKLSGCPDLDFINONNIPMYLMF 840
Qy 841 DSLNNSLSQLEFMEIYKEMVCNLYKLDNDKIKNLLEBAKRVSTSVKTLSSSSMPLSLT 900
Db 841 DSLNNSLSQLEFMEIYKEMVCNLYKLDNDKIKNLLEBAKRVSTSVKTLSSSSMPLSLT 900
Qy 901 PODPREVANDTSHSTNLNSIKLFENILSLGKKNITYOELIGKSSSENYEKLKXSD 960
Db 901 PODPREVANDTSHSTNLNSIKLFENILSLGKKNITYOELIGKSSSENYEKLKXSD 960
Qy 961 TFYNESEFNFYKSKRADINSINDESKRKLEEDINKLKTQJLSFDLYNKKIKLERLFD 1020
Db 961 TFYNESEFNFYKSKRADINSINDESKRKLEEDINKLKTQJLSFDLYNKKIKLERLFD 1020
Qy 1021 KKTIVGKYMOKIKLUTLLKEQLESKLSLNNPKHVLQNFVFNKKKBAEIAETENTLEN 1080
Db 1021 KKTIVGKYMOKIKLUTLLKEQLESKLSLNNPKHVLQNFVFNKKKBAEIAETENTLEN 1080
Qy 1081 TKILLLKHKGLVKKYNGSSPLKLTISESIOEDNYASLENFKYLSKLEGLKDNLEK 1140
Db 1081 TKILLLKHKGLVKKYNGSSPLKLTISESIOEDNYASLENFKYLSKLEGLKDNLEK 1140
Qy 1141 KKLVSLSGGLHLLAELKEVILKKNKYTNSPSENNTDVNNALESKPKPLPGTDAVATVS 1200
Db 1141 KKLVSLSGGLHLLAELKEVILKKNKYTNSPSENNTDVNNALESKPKPLPGTDAVATVS 1200
Qy 1201 ESGSDTLEQSQPKRPASTHVAESNTITTSQNVDEVDVILVPLFGESEBDYDGLGVV 1260
Db 1201 ESGSDTLEQSQPKRPASTHVAESNTITTSQNVDEVDVILVPLFGESEBDYDGLGVV 1260
Qy 1261 TGEAVTPSVINILSKISKENIEVLYLKLAVYYSKLKOLENNWTFVANNKDIINSRPN 1320
Db 1261 TGEAVTPSVINILSKISKENIEVLYLKLAVYYSKLKOLENNWTFVANNKDIINSRPN 1320
Qy 1321 KREPNKNVLESDDLKYKOLTSNNVVDYKPYKFLNKEKBDKFLSSNNYIKDISIDPIDINAN 1380
Db 1321 KREPNKNVLESDDLKYKOLTSNNVVDYKPYKFLNKEKBDKFLSSNNYIKDISIDPIDINAN 1380
Qy 1381 DVLGYKILSEKYSDDLSIKKYINDKOGENEKYLPLFNINIEITYKYTVNDKIDL.FVJHLE 1440

Db 1381 DVLGYKILSEKYSDDLSIKKYINDKOGENEKYLPLFNINIEITYKYTVNDKIDL.FVJHLE 1440
Qy 1441 AKVLNYYTEKSNVEKIKELVLYKLTIOQKLDKFKKNNNFVGIADLSDYNNNNLLTKFLS 1500
Db 1441 AKVLNYYTEKSNVEKIKELVLYKLTIOQKLDKFKKNNNFVGIADLSDYNNNNLLTKFLS 1500
Qy 1501 TGMVFNIAKTIVLSNLDGNGMLNISQHOQVKKOCQPNSGCFRHLDERECKCLLANK 1560
Db 1501 TGMVFNIAKTIVLSNLDGNGMLNISQHOQVKKOCQPNSGCFRHLDERECKCLLANK 1560
Qy 1561 QEGDKCVENPNPTCENNNGCCDADAKCTEEDSGSNKKITCECTKPDSPPLPDGIFCSSS 1620
Db 1561 QEGDKCVENPNPTCENNNGCCDADAKCTEEDSGSNKKITCECTKPDSPPLPDGIFCSSS 1620
Qy 1621 NFGISFLIIMLILYSFI 1639
Db 1621 NFGISFLIIMLILYSFI 1639

RESULT 2
A24594
probable major surface antigen (83K, 19K, 42K) precursor - malaria parasite (Plasmodi
C:Species: Plasmodium falciparum
C:Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 09-Jun-2000
C:Accession: A24594
R:Holder, A.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholl
Nature 317, 270-273, 1985
A:Title: Primary structure of the precursor to the three major surface antigens of Pl
A:Reference number: A24594; MUID:86014355
A:Molecule type: DNA
A:Residues: 1-1640 <HOL>
C:Superfamily: 1-majozeroite surface antigen
C:Keywords: surface antigen

Query Match 97.2%; Score 8191.5; DB 2; Length 1640;
Best Local Similarity 97.9%; Pred. No. 9.8e-247;
Matches 1605; Conservative 8; Mismatches 26; Indels 1; Gaps 1;

Qy 1 KMIIFELCSLPIFIINTOCVTHESYOELVKKLEDAEDAVLTGYSLFQEKKNVLENGTSGT 60
Db 1 KMIIFELCSLPIFIINTOCVTHESYOELVKKLEDAEDAVLTGYSLFQEKKNVLENGTSGT 60
Qy 61 AVTSTPGSKGSVASGSGSVASGGSVASGGSVASGGSVSGSNRRTPSPDSSDS 120
Db 61 AVTSTPGSKGSVASGSGSVASGGSVASGGSVASGGSVSGSNRRTPSPDSSDS 120
Qy 121 DAKSYADLKHVRNVLTLTKELKYPQLFDLTNNHMLTLCDNTHGRKYLLIDGKEEINELLK 180
Db 121 DAKSYADLKHVRNVLTLTKELKYPQLFDLTNNHMLTLCDNTHGRKYLLIDGKEEINELLK 180
Qy 181 LNFYFDLIRAKLNDVCANDYQOIPPNLKITRANEDVLLKLVFGYRKPLDNIKDNVGMED 240
Db 181 LNFYFDLIRAKLNDVCANDYQOIPPNLKITRANEDVLLKLVFGYRKPLDNIKDNVGMED 240
Qy 241 YIKKKNKTIENINELIESKKTIDKKNKATKEEEKKKIYQAOYDLSYNNKOLEFAHNLS 300
Db 241 YIKKKNKTIENINELIESKKTIDKKNKATKEEEKKKIYQAOYDLSYNNKOLEFAHNLS 300
Qy 301 VLEKRIIDLTKNENIKELLDKINEIKNPPANSNGTPTLLDKNKIEHEKEIKETAKT 360
Db 301 VLEKRIIDLTKNENIKELLDKINEIKNPPANSNGTPTLLDKNKIEHEKEIKETAKT 360
Qy 361 IKFNIDS.LFTDPLELEYLREKKNKIDISAKVETKESTEPENYENGVTYPLSYNDINNAL 420
Db 361 IKFNIDS.LFTDPLELEYLREKKNKIDISAKVETKESTEPENYENGVTYPLSYNDINNAL 420
Qy 421 NELNSFGDLINPFDYTKPSKNITTDNERKKFINEIKKIKIEKKKIESDKSYEDRSKS 480
Db 421 NELNSFGDLINPFDYTKPSKNITTDNERKKFINEIKKIKIEKKKIESDKSYEDRSKS 480

QY	481	LNDITKEVKEILLNETYOSKFNANNIDLTNEFKAMGKRXYKVEKLTJHNTFASYENSKNL	540
Dd	481	LNDITKEVKEILLNETYOSKFNANNIDLTNEFKAMGKRXYKVEKLTJHNTFASYENSKNL	540
QY	541	EKLTKALTYMEDYSLRNIYVEKELKYKNNLSIKIENIEFTLVANIKKDEEOLPEKKTID	600
Dd	541	EKLTKALTYMEDYSLRNIYVEKELKYKNNLSIKIENIEFTLVANIKKDEEOLPEKKTID	600
QY	601	ENKPEDEKILEVSDIVKVOVKVLLMNKIDELAKFTOILLKAVELKANIHPNYSKOENKO	660
Dd	601	ENKPEDEKILEVSDIVKVOVKVLLMNKIDELAKFTOILLKAVELKANIHPNYSKOENKO	660
QY	661	PYYLIVLKEEIDKLKVPMPVVESTINEEKNIKITGOSDSNESEPTGCEITTGATTKPGQO	720
Dd	661	PYYLIVLKEEIDKLKVPMPVVESTINEEKNIKITGOSDSNESEPTGCEITTGATTKPGQO	720
QY	721	AGSALGSGSYOAOOEKOQOAPVPVPVPAKQOVPPAPVANNKRENSKIDYLEKLYE	780
Dd	721	AGSALGSGSYOAOOEKOQOAPVPVPVPAKQOVPPAPVANNKRENSKIDYLEKLYE	780
QY	781	FLNLSYICHKIYLVSHSTMEKILKQYKIRKEEESKLSGCDPDLFNFJONNIPWYSMF	840
Dd	781	FLNLSYICHKIYLVSHSTMEKILKQYKIRKEEESKLSGCDPDLFNFJONNIPWYSMF	840
QY	841	DSLNNLSQLEFMEIYERKEMVONLYKLDKNDKIKNLLEAKKYSTSVKTLSSSSMPLST	900
Dd	841	DSLNNLSQLEFMEIYERKEMVONLYKLDKNDKIKNLLEAKKYSTSVKTLSSSSMPLST	900
QY	901	PODKPEVSANDTSHSNLNNLSKLFPNLSLCKKNKIYOEILGOKSSSENYEKILKOSD	960
Dd	901	PODKPEVSANDTSHSNLNNLSKLFPNLSLCKKNKIYOEILGOKSSSENYEKILKOSD	960
QY	961	TFYNESFPNEFKSADDTINSJNDESKRKLEEDINLKLTLOLSPLYKRYKTLKLEPLD	1020
Dd	961	TFYNESFPNEFKSADDTINSJNDESKRKLEEDINLKLTLOLSPLYKRYKTLKLEPLD	1020
QY	1021	KKTGVKYYKMOIKKTLTKLEQESKLNSLNNPKHYLONTSVEFNKKKEALAEETENTLEN	1080
Dd	1021	KKTGVKYYKMOIKKTLTKLEQESKLNSLNNPKHYLONTSVEFNKKKEALAEETENTLEN	1080
QY	1081	TKIILKHKGGVKKYVYNNESSPLKTLSESIQOTEDNVAASLENKRYLSKLEGKLKONLNEK	1140
Dd	1081	TKIILKHKGGVKKYVYNNESSPLKTLSESIQOTEDNVAASLENKRYLSKLEGKLKONLNEK	1140
QY	1141	KKLSTLSSGLHLHLAELEKVIYKKNKNYTGNSPENNTDVNNALESYKFPJPEGTDVAITYS	1200
Dd	1141	KKLSTLSSGLHLHLAELEKVIYKKNKNYTGNSPENNTDVNNALESYKFPJPEGTDVAITYS	1200
QY	1201	EGSGDPLFEOQPKKPASTHVGAESNNTITTSQNVDEVDVYIIVPJFGESEEDYDILGOVY	1260
Dd	1201	EGSGDPLFEOQPKKPASTHVGAESNNTITTSQNVDEVDVYIIVPJFGESEEDYDILGOVY	1260
QY	1261	TGGEAVTSGVIDNLSKIENEYEVLYLKLPLAGVYRSJKOLENNVMTFNVNVKDILNSRFN	1320
Dd	1261	TGGEAVTSGVIDNLSKIENEYEVLYLKLPLAGVYRSJKOLENNVMTFNVNVKDILNSRFN	1320
QY	1321	KRENFKNVLESDDLPRYKODLTSSNYYKADPKFPLNKEKROKPLSSYNYIKDSIDPTJINFRAN	1380
Dd	1321	KRENFKNVLESDDLPRYKODLTSSNYYKADPKFPLNKEKROKPLSSYNYIKDSIDPTJINFRAN	1380
QY	1381	DVLGYKYLSEKKYSSDLSIKKYINDKOGENEKYLPFLNNIEFTLYKVNDKIDLFPVHLE	1440
Dd	1381	DVLGYKYLSEKKYSSDLSIKKYINDKOGENEKYLPFLNNIEFTLYKVNDKIDLFPVHLE	1440
QY	1441	AKVLYNTRYKSNVEVKIKELNYLTKTIQDKLADPKKNNNVGJADISTDYNNHNNLTKFLS	1500
Dd	1441	AKVLYNTRYKSNVEVKIKELNYLTKTIQDKLADPKKNNNVGJADISTDYNNHNNLTKFLS	1500
QY	1501	TGMVFENLAKTYVSNLNDJGMNLGMINISOHCVKQOCPONSGCFPHIDERECEKCLANY	1559
Dd	1501	TGMVFENLAKTYVSNLNDJGMNLGMINISOHCVKQOCPONSGCFPHIDERECEKCLANY	1559
QY	1560	KQEBDCVENDNPNTCENNNGCDAADAKCTEEDSGSNKKRTGECTKPDSPYLPFDGIFCSS	1619

```

Db      1561 KOEGSCVENSPTCENNENGGCDADAKCTEEDSGSNGKRKICECTPCDPLSMWIFCSS 1620
OY      1620 SNFLGISFLLIMLILYSFT 1639
Db      1621 SNFLGISFLLIMLILYSFT 1640

RESULT      3
SAZOKI
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 30-Sep-1987 #sequenceRevision 30-Sep-1987 #textChange 09-Jun-2000
C:Accession: A25120
R:MacKay, M.; Goman, M.; Bore, N.; Hye, J.E.; Scalfe, J.; Certa, U.; Stunnenberg, H.
EMBO J. 4, 3823-3829, 1985
A>Title: Polymorphism of the precursor for the major surface antigens of Plasmodium f
A:Reference number: A91030; MUID:86136024
A:Accession: A25120
A:Molecule type: DNA
A:Residues: 1-1631 <MAC>
C:Comment: The merozoite stages of different strains have strain-specific surface ant
C:Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The
C:Superfamily: Major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat; transmembrane pr
F:1-19/Domains: signal sequence #status predicted <SIG>
F:20-1631/Product: major merozoite surface antigen #status predicted <MAT>
F:67-84/Region: 3-residue repeats (S-G-T/P)
F:1614-1631/Domains: membrane anchor #status predicted <MBN>
F:97,259,755,759,835,911,955,1049,1156,1165,1436,1563/Binding site: carbohydrate (Asn

Query Match          94.6%; Score 7965; DB 1; Length 1631;
Best Local Similarity 95.3%; Pred. No. 1e-239;
Matches 1563; Conservative 19; Mismatches 48; Indels 10; Gaps 3;

OY      1 MKIIIFLCSEFFIINTGCVTHESYQELVKLTLEALEDAVLGYSLFOEKMYLNESGTSGT 60
Db      1 MKIIIFLCSEFFIINTGCVTHESYQELVKLTLEALEDAVLGYSLFHKKMKILNEBEETTT 60
OY      61 AVTTSTPGSKGSVASGGSGSVASGGSVASGGSVASGGSVASGGSNSRTPNSDSDDS 120
Db      61 KGASAQSQTSGT--SGTSGPSGSPGTSPPSSRSNTLPRTSNTSGAS-----PADASDS 111
OY      121 DAKSVADLKHRVRNRLTLTKELKYPOLFDNLNMHTLCDNHGRKYLIDGYEEINELLTX 180
Db      112 DAKSVADLKHRVRNRLTLTKELKYPOLFDNLNVHLITCDNHGRKYLIDGYEEINELLTX 171
OY      181 INFNYDDLRAKLNDVCANDYCQIPFNLIKIRANELDVLKLVFGYRRPLDNKDYNVGKMD 240
Db      172 INFNYDDLRAKLNVNCANDYCQIPFNLIKIRANEIDLVLKLVFGYRRPLDNKDYNVGKMD 231
OY      241 YIKNNKKTIEINIELIEESKKTIDKNNAATKEEEKKKLYQAQYDLSTYNKOLEEAHNLIS 300
Db      232 YIKNNKKTIEINIELIEESKKTIDKNNAATEEBEKKKLYQAQYDLSTYNKOLEEAHNLIS 291
OY      301 VLEKRIDTLKKNENIKELDKINEIKNPFPANSQTPPTLLDDKKKIIEEHKEKETELAKT 360
Db      292 VLEKRIDTLKKNENIKELDKINEIKNPFPANSQTPPTLLDDKKKIIEEHKEKETELAKT 351
OY      361 IKFNIDSFTPTPLELEYLRERKNKNIDISAVENFKESTEPENYGTYPLSYNDINNAL 420
Db      352 IKFNIDSFTPTPLELEYLRERKNKNIDISAVEFKESTEPENYGTYPLSYNDINNAL 411
OY      421 NELNSFGGLINPEDYTKEPSKNITYTDNERKKFIWEIKEKIKIEKKKIESDKSYEDRSKS 480
Db      412 NELNSFGGLINPEDYTKEPSKNITYTDNERKKFIWEIKEKIKIEKKKIESDKSYEDRSKS 471
OY      481 LNDITRKVEKLLNELNYOSKPNNNIDLTNEEKMMKRRSYKEYKLTJHNTPASYNSKHNL 540
Db      472 LNDITRKVEKLLNELNYOSKPNNNIDLTNEEKMMKRRSYKEYKLTJHNTPASYNSKHNL 531
OY      541 EKLTKALYMEDYSLRNIIVVERKEELKYKNLSIKLENEIETLTVEVNIKKDEDEDLFEKKTKTD 600

```

Db	532	EKLTKALYEMEDYSLRNIIVYELKLYKNIISLENIEMIFIVENIKOEBOJFEKKITIKD	591
Qy	601	EKKPDEKLEIVSDIVKYOVOKVYLLANKIDELKKTQULIKNVELKNIHVPNSYKOEKOE	660
Db	592	EKKPDEKILEVSDIVKYOVOKVYLLMNKIDELKKTQULIKNVELKNIHVPNSYKOEKOE	651
Qy	661	PYYLIVLKELEIDKLYMPKVESIINEKKKNITIEGOSNSEPTEGTELTOGATKPCQO	720
Db	652	PYYLIVLKELEIDKLYMPKVESIINEKKKNITIEGOSNSEPSTEGETTGATTKPCQO	711
Qy	721	AGSALEGDSVOAQOEKOQOPVPVPVPEPAKAQVPTPAPYNNKTENVSKIDYLEKYE	780
Db	712	AGSALEGDSVOAQOEKOQOPVPVPVPEPAKAQVPTPAPYNNKTENVSKIDYLEKYE	771
Qy	781	PLNTSYIOHKIILVSHSTMEKELIKQYKTKEEBSKSSCDPLDLFNIONNIPVMSMF	840
Db	772	PLNTSYIOHKIILVSHSTMEKELIKQYKTKEEBSKSSCDPLDLFNIONNIPVMSMF	831
Qy	841	DSLNISSLQLEMEIYEEKEMWCNLYKLIKDNOKIKNLEAKKVTSTVKTITSSSMOPLST	900
Db	832	DSLNISSLQLEMEIYEEKEMWCNLYKLIKDNOKIKNLEAKKVTSTVKTITSSSMOPLST	891
Qy	901	PODKREVSANDTSHSTNLNLSKLEPENILSLCKNNKIIOELIGKSSENEPEKILKDS	960
Db	892	PODKREVSANDTSHSTNLNLSKLEPENILSLCKNNKIIOELIGKSSENEPEKILKDS	951
Qy	961	TFYNSFPNFKSKADDINSINDESKRKKLEBDINKLKTQLQSPDLVYKFKLERLFD	1020
Db	952	TFYNSFPNFKSKADDINSINDESKRKKLEBDINKLKTQLQSPDLVYKFKLERLFD	1011
Qy	1021	KKKTYGKKMOIKKLTLLKLEBSKLSNLNPNPHVLONSVPEFNKKKEATIAEPTENTEN	1080
Db	1012	KKKTYGKKMOIKKLTLLKLEBSKLSNLNPNPHVLONSVPEFNKKKEATIAEPTENTEN	1071
Qy	1081	TKILKHKYKGLVKKYKNGESSPLTISESIQOTDNYASLENKRVLSKLEGLKLDNLNEK	1140
Db	1072	TKILKHKYKGLVKKYKNGESSPLTISESIQOTDNYASLENKRVLSKLEGLKLDNLNEK	1131
Qy	1141	KKLSTLSSGLHHLIAELKEVIKKNKNTGNSPSENNTDVNNALESYKKFLPECTGVATVS	1200
Db	1132	KKLSTLSSGLHHLIAELKEVIKKNKNTGNSPSENNTDVNNALESYKKFLPECTGVATVS	1191
Qy	1201	EGSGOTLEOQPKPASTHNGAESNNTTTSQNDDEVDVYIIVPFGSEBBDYDLGVY	1260
Db	1192	EGSGOTLEOQPKPASTHNGAESNNTTTSQNDDEVDVYIIVPFGSEBBDYDLGVY	1251
Qy	1261	TGEAVTESVIDNILSKLENEYEVLVYKPLAGVYRSKLOLENNWMTFVNNVKDINSRBN	1320
Db	1252	TGEAVTESVIDNILSKLENEYEVLVYKPLAGVYRSKLOLENNWMTFVNNVKDINSRBN	1311
Qy	1321	KRENFKNVLESDLIPYDLTSSNVVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFN	1380
Db	1312	KRENFKNVLESDLIPYDLTSSNVVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFN	1371
Qy	1381	DVLGYYKILSEKRYKSDLSIKKIYNKOGENEKYLPEFLNNITELTKYVNDKIDLFVHLE	1440
Db	1372	DVLGYYKILSEKRYKSDLSIKKIYNKOGENEKYLPEFLNNITELTKYVNDKIDLFVHLE	1431
Qy	1441	AKVLYNTYEEKSNEVVKLEKNYLTKTQDKLADPKKNNNVGLADISTDYNNHNLITKFLS	1500
Db	1432	AKVLYNTYEEKSNEVVKLEKNYLTKTQDKLADPKKNNNVGLADISTDYNNHNLITKFLS	1491
Qy	1501	TGAVFENLAKTVLSNLDGNUL-OGMILNISOHCVKKQOCFQNSGCFERHLDERECKCLNTY	1559
Db	1492	TGAVFENLAKTVLSNLDGNUL-OGMILNISOHCVKKQOCFQNSGCFERHLDERECKCLNTY	1551
Qy	1560	KOEBDKCVENBPPTCENNNGCDAADAKCTEBDSGNGKKTCECTKPDSPYPLFDGIFCSS	1619
Db	1552	KOEBSKCVENBPPTCENNNGCDAADAKCTEBDSGNGKKTCECTKPDSPYPLFDGIFCSS	1611
Qy	1620	SNFLGISFLILMLILYSFI1639	

DB 1612 SNFGLISFLLMLMLVYSFI 1631

RESULT 4

SAZQOM

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)

N:Alternate names: 195K glycoprotein

C:Species: Plasmodium falciparum

C:Date: 30-Sep-1987 #sequence_revision 31-Mar-1991 #text_change 09-Jun-2000

C:Accession: A23386; S06361

R:Weber, J.L., Leininger, W.M., Lyon, J.A.

Nucleic Acids Res. 14, 3311-3323, 1986

A>Title: Variation in the gene encoding a major merozoite surface antigen of the human

A:Reference number: A23386; MUID:86205236

A:Accession: A23386

A:Molecule type: DNA

A:Residues: 1-1104 <WEB1>

A:Cross-references: EMBL:X03831

R:Weber, J.L., Sim, B.K.L., Lyon, J.A., Wolff, R.

Nucleic Acids Res. 16, 1206, 1988

A>Title: Merozoite surface protein sequence from the Camp strain of the human m

A:Reference number: S06361; MUID:88143999

A:Accession: S06361

A:Molecule type: DNA

A:Residues: 1104-1726 <WEB2>

A:Cross-references: EMBL:X03831

C:Comment: The merozoite stages of different strains have strain-specific surface ant

C:Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The

C:Superfamily: major merozoite surface antigen

C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-1726/Product: major merozoite surface antigen #status predicted <MAT>

F:67-87, 91-96, 100-105, 109-120/Region: 3-residue repeats (S-G-T)

F:757-765/Region: 3-residue repeats (T-E-E)

F:133, 212, 501, 567, 638, 827, 839, 924, 944, 990, 1016, 1114, 1221, 1613, 1658/Binding site: carb

Query Match	61.0%;	Score 5136.5;	DB 1;	Length 1726;
Best Local Similarity	60.1%;	Pred. No. 4.1e-152;		
Matches 1052;	Conservative 221;	Mismatches 342;	Indels 135;	Gaps 26;

[illegible]

Qy	446	DN-ERKFNIEIEKIKIEKKI-----EDKSYEDRSJLMDITKEVEKLLNIEYDSKF	500
Db	478	DNKREKLFINNKKQIDLEKKIKNTHREONKKLEBYERS----KKDYELIEKKEKFF	533
Qy	501	NNNIDLNFPEKMGKRSRYVEKLTTHNFPASYENSKNHLKLTFKALYMEDYSLRNTIV	560
Db	534	NNNDRKDVOKISARTYVKEKORINKNKFFSSNNSVYVQKLLKALSLYLEDYSIRKIS	593
Qy	561	EKELKYKNIJSKIENEIETLVENIKKDEOULFEKK--ITKDEKNPDEKILLEVSDIKV	617
Db	594	EKDENNHYTLKTGLHADIKKLEPEIKSSENKILKFNKGLTHSANA-----LEVYDIKL	649
Qy	618	QVQVAVLLMANKIDELKTQULIKVAVELKNHINVPNSYKOEKOEPRYLLVYLKKEIDKLVF	677
Db	650	QVQVAVLLIKRIEDLRKLETFELKMAOLKDSIHVPNITKPPKPEPRYLLVYLKKEVDKKEF	709
Qy	678	MPKYESLINEKKNIKEEGSDNSEPSTGEITQATTKPQGOQASALJEGSOVQAQAOEQ	737
Db	710	IPKXMDLKKQEAVALS-----ITQPLVAASFTTDEGSHHTLSQSGETEVEYTEP	760
Qy	738	KQAOAPV-----RPVPEAKAQVPTPEAPVNNNTENWASKLDYLEKLEFELNTSYCH	789
Db	761	EETBEVYGHNTTYITILPREVAVVENSLEHKSNDOSQALMTYVLLKDKLEDTFYSGYCH	820
Qy	790	KYIIVSHSTANEKILQOYKITYKEESKLSLSCDPLDLFNIONNIPVMTSMEDSLNLSISQ	849
Db	821	KYIIVSSNSMDOKILEVYNLTPEEENELSCDPLDLFNIONNIPAMSYLSDSMNDIQH	880
Qy	850	LFMEITYKEWVCNLYTKDKNDIKNLEFKKV-----SPSYVTLSSS-----	893
Db	881	LFEPLOYKEMITYUHKIKENHNHILKLEEROKOITGTSSTSPGNTTVTAQSAATHSNQ	940
Qy	894	--MOPSLTPQDQREVSAND---DTSHS-----TNLNSILKLEPNLISLGNKNITYOEL-I	943
Db	941	QOSMASTMTNQNGAAVSSGVAVEESHDPDLVYLSINDLKGIVSLNIGNKTKVPNPLTI	1000
Qy	944	GQKSENFYEKILKSDSTFYNESFTFNVKSKADINSLNDESKRKKLEBDINKLKTUOL	1003
Db	1001	STTEMEFYEENILKNDTYTENDDIKOFVSNKSXVITGLT-ETOKNALNDEIKKIKLDTUOL	1059
Qy	1004	SFDLYNNKYKILKLEBILPKKKTVGKYMOJKLTLKLEOLSKLNSLNPKNHLYONFSVF	1063
Db	1060	SFDLYNNKYKILKLEBILPKKKLEQDMQKIKLTLKLEQESKLSLNSNPNHLYONFSVF	1119
Qy	1064	NKKKEAEIAETENMLENTKTLKLNHYKGLVKKYUGESSPLKTLSEESIJOTEDYASLENFK	1123
Db	1120	NKKKEAEIAETENMLENTKTLKLNHYKGLVKKYUGESSPLKTLSEYSIJOTEDYANLEKFR	1179
Qy	1124	VLSKLESKLDNINLEKKKLSYLSGLJHLHLAEKLEVINKNKYTGNSPSENNTDVNNALE	1183
Db	1180	VLSKIDGKLDNNDHLGKRRKLSFLSSGJHLHLTELKEVINKKMYTGNSPSENNNKRVNALK	1239
Qy	1184	SYKFFLPBEGDVFATVSE-----SGSOTLEQSPKKPASHYGAESNTITTT	1229
Db	1240	SYENFLEP-AKVTTVVTPPOPDVTPSPLSVRVSGSSGSKTEQTOIPTSGSLTELEQVOVQ	1298
Qy	1230	SONYDEVDVYIYIPJIGSESEEDYDGLQVYGEAVTPSVININISKIENEYEVYLP	1289
Db	1239	LQNYDEDDDSLYVLPJIGSEEDNDVLYQVYGEALSYT-MONILSGPENEVDYIYLP	1357
Qy	1290	AGVYRSILKLOLENNVMTFVNVKQILNSRPNKREKNVLESDDLIPYKDLTSSNVYKDP	1349
Db	1358	AGVYRSILKLOLENNVMTFVNLNDINSLRKRRKYPFLVLESDDLQMFKHSISNEYIEDS	1417
Qy	1350	YKFLNKKERBOKFLSSVNYIKDSITDQINPANDVLGYKYLKLSKYSOJDSIKKYI-----	1404
Db	1418	FKLLNSQKNTLKSXYKIKESVENDIKFAQOEISYEVKLVAKYKDDLESIKKVIKEKE	1477
Qy	1405	-----NDKQSENKYLPFLNNIETTKYANDIDTIDFVILIKELVANTYVE	1449
Db	1478	KFPSSPPTTPPSAKTDEQKREKFLPEPLNITETLNNLVKNKIDYILNKAKKINDCAVE	1537
Qy	1450	KSNVEVAKIKELNLTKTQDKLADFKKNNNFVGIADLSTDYNNHNLITKFLSTGVAFENILA	1509

[illegible]

```

Db 478 DNKEKRIEINNKKOIDEKKINHTEQONKKLLEDYERS---KKDYEELEKEFEYENK 533
Qy 501 NNNDILTEFKMKQRYKYVEKLTHTNHTFASYENSKHNLEKTLKALYMEDYSLRNIV 560
Db 534 NNNDKDVVDKIFISARYYNNKORNNKPFSSNNNSYNNVOKLAKALSYLEDYLSRKIS 593
Qy 561 EKEKYYKNLISKIENELETLEVENIKKDEQLFEKK---ITKDNKPKDEKILEVSDIVK 617
Db 594 EKDNNHYETLKTGLADIKLLEEKISENNKILEKFNKGLTHSANAS---LEVYDIKVL 649
Qy 618 QOVAVLLMKNKIDELKTKQOLIKNVELKHNHVPNSYKQDNQOEPRYLLVKKEDKLKVF 677
Db 650 QOVAVLLIKKIEDIKLEFLKMAQLKDSIHVPNLYKPPONKPEPYLLVLEKVDKLEF 709
Qy 678 MPKESLITNEKKNKITEGQSDNSEPTEGELTGQATKPPQOAGSALEGDSVQAQOQ 737
Db 710 IPKPKMKKQOAVLS-----ITQPLVASETTEDGSHNLTLSOGTEFEVTEET 760
Qy 738 KOAOPV-----PVPVPEAKAOPPPAPVNNKTEVNSKLDYLEKLEFLNTSYICH 789
Db 761 EETEBVTGHTTIVTTLPRKVKYVENSIKHSNDNSQALRTKYVLKLDLEFLTSYICH 820
Qy 790 KYIIVSHSTNMEKLIKOKYKITEESKLSGCDPLDLFNIONNIPVMYSMFDLSNLSIQ 849
Db 821 KYIIVSNSSMOQKLEEVNLTPEEENELKSCDPLDLFNIONNIPAMYSLYDSMNNDLQH 880
Qy 850 LFMETYEKEMCNLYKLDNDKIKRLBEAKKV-----STVYKTLSSSS----- 893
Db 881 LFFELYQKEMYYIKLKEENHIKLEDEQKOITGSTSTSPGNTTVAOSATHSNQON 940
Qy 894 --MOPSLTPDKREVSAND---DTSHS---TNLNSLKLEFNILSLGKNKNYQEL-I 943
Db 941 QOSNASTNTONGVAVSSGPAVVEESHDPILVLTISNDLKGIVSLNNGKTKVNPPLTI 1000
Qy 944 GOKSENFYEKILDSDFEYSEFNPKSKADINDSLNDSKRRKLEEDINLKLTLO 1003
Db 1001 STTEMERYEYILKNNDYFENDDIKQFVKSNSKYITGLT-ETQKALNDEKIKKLTLO 1059
Qy 1004 SFDLYNKKYKLERLEFDCKRTVGYKMOIKKLTLLKEOLESKLNSLNPKVYLONFVFF 1063
Db 1060 SFDLYNKKYKLERLEFDCKRTVGYKMOIKKLTLLKEOLESKLNSLNPKVYLONFVFF 1119
Qy 1064 NKKKEAEIETENTLENTKILKHYKGLVKKYNGESSPLKTLSESIOTEDNYASLENK 1123
Db 1120 NKKKEAEIETENTLENTKILKHYKGLVKKYNGESSPLKTLSESIOTEDNYANLEKFR 1179
Qy 1124 VLSLLEGKLNKLNLEKKKLSYSSGLHLLAELKEVYKKNKNYNGSSNNKTVNNALE 1183
Db 1180 VLSKIDGLNLDNLHGGKKLSFSSGLHQLITLEKEVYKKNKNYNGSSNNKTVNNALE 1239
Qy 1184 SYKFFLEPGTVAIVVSE-----SGSDTLEQSQPKKPASTHGAESNTITTT 1229
Db 1240 SYENFLPE-AKVTVTVVPPQDVTVPSPLSVAVSSGSGSTKETQIPTSGLTLEQOYVQ 1298
Qy 1230 SQNVADVEDVIVIPFGESEEDYDGLQVVTGAVTPSVIDNLSKILENEEYVLYKPL 1289
Db 1299 LQNYDEEDDSLVVLPFGESEEDNDEYLDQVVTGAVT-NDNLSGSENEEYVLYKPL 1357
Qy 1290 AGVYRSJLKOLENNVMTFNVAVKDLNSRFKKNREKFAVLESOLIPYKDLTSSNYVAKDP 1349
Db 1358 AGVYRSJLKOLENNVMTFNVAVKDLNSRFKKNREKFAVLESOLIPYKDLTSSNYVAKDP 1417
Qy 1350 YKFNKREKROFELSYYIKSDIDINFANDVGYKYLSEKYSKSDSISKKYI----- 1404
Db 1418 FKLNSQKNTLKSYYIKSIVENDIKFAQEGISYIKVYLAKKDLDESIAKKYIKEKE 1477
Qy 1405 -----NDQGENEKYLLPLNIEFLYKTVNDKIDLEFVTHLEKAVLYNTE 1449
Db 1478 KFPSSPTTPPSPVKTDEQKESKFLPLTNIEFLYNNLVKNDYILNLAKINDCIVE 1537
Qy 1450 KSNVEVATKELNLYKTODKLADFKKNNNPGVADLSIDYVHNNLLKFLSTGAVFELA 1509
Db 1538 KDEAHVATITKLSIDKADKIDLFKNHNDFAIKLINDTKKMDLGLSTGLV-QNFP 1596

```

```

Qy 1510 KTVLSNLDGNLQGMNLNISQHCYKQOPNSGCFRHLDEBECKCLNTYKQEDKCVEN 1569
Db 1597 NTIISKILEGKFOMLNLSQHCYKQOPNSGCFRHLDEBECKCLNTYKQEDKCVEN 1656
Qy 1570 PNPTCENNNGGCDADAKTEEDSGSNGKKTCECTKPDSPYLPFGJICSSNFGISFLL 1629
Db 1657 PNPTCENNNGGCDADAKTEEDSGSNGKKTCECTKPDSPYLPFGJICSSNFGISFLL 1716
Qy 1630 LTMILYVSI 1639
Db 1717 LTMILYVSI 1726

```

RESULT 6

```

A:54498
major: merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 28-Oct-1994 #sequence: revision 28-Oct-1994 #text_change 09-Jun-2000
C:Accession: A54498
R:Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown
Mol. Biochem. Parasitol. 27, 291-302, 1988
A:Title: Variation in the precursor to the major merozoite surface antigens of Plasm
A:Reference number: A54498; MUID:88142999
A:Accession: A54498
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1701 <PEP>
A:Cross-references: GB:M19143; NID:g160412; PIDN:AAA29653.1; PID:g160413
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

```

```

Query Match 60.1%; Score 5065; DB 2; Length 1701;
Best Local Similarity 60.1%; Pred. No. 6; 6e-150;
Matches 1042; Conservative 227; Mismatches 338; Indels 126; Gaps 29;

```

```

Qy 1 MKIIFLCSPFLFTINQCVTHESYQELVKLEALEDAVLVGYSLFQEKKNVLEGTSGT 60
Db 1 MKIIFLCSPFLFTINQCVTHESYQELVKLEALEDAVLVGYSLFQEKKNVLEGTSGT 60
Qy 61 AVTSTPGSGSVASGSGSVASGSGSVASGSGSVASGSGSGNSRTPNSDSSS 120
Db 61 AVTSTPGSGSVASGSGSVASGSGSVASGSGSGNSRTPNSDSSS 114
Qy 121 DAKSYADLKHVRNLYLTIKELKYPOLFDLTNHLTLTCDNHHGKRYLIDGEEFINELLYK 180
Db 115 NFKTYADLKHVRNLYLTIKELKYPOLFDLTNHLTLTCDNHHGKRYLIDGEEFINELLYK 174
Qy 181 LNFYFDLLRAKLNVCANDYQIFPNLIRANELDVLKLVFGYRKPDLNIDKQVGMED 240
Db 175 LNFYFDLLRAKLNVCANDYQIFPNLIRANELDVLKLVFGYRKPDLNIDKQVGMED 234
Qy 241 YIKKNKTTIENINLEESKKTIDKNKNATKEEKKKLYQAQYDLSTYNNQLEAHNLIS 300
Db 235 YIKKNKTTIENINLEESKKTIDQNNKADNEEGKKLYQAQYDLSTYNNQLOEAHNLIS 294
Qy 301 VLEKRIDTLLKKNENIKELDLKINIK-----NPPRANGNPNTLLD-KNKKIEHEKEIK 355
Db 295 VLEKRIDTLLKKNENIKELDLKIDKIDAEHP--TTGSKRNPPLPENKKKEVEHEBEKIK 351
Qy 356 EIATIKFINIDSLFTDPLELEYLYREKKNKIDISAKVE--TKESTEEN-EYKPNVTPPLS 412
Db 352 EIATIKFINIDSLFTDPLELEYLYREKKNKVDYVPKQDPFKSQVQIFKVPNPIVPLP 411
Qy 413 YNDINNAL--NELNSFGDLINPFDYTKEPSKNITYDN-ERKKEFINIEKIKIEKKKI- 467
Db 412 LFDIHNLSLADNDKNSYGDLMNP-DTKEXINEKILITDNKERRKIFINNKKOIDELEKRN 470
Qy 468 ---ESDKKSYEDRSKSLNDITKEYEKLLNELYDQSKFNNDIDLNFEEKMGRKSYKEKL 524
Db 471 HFKQONKKLLEDYERS---KKDYEELEKEFEYENKFPNNNDKDVVDKIFASRYTYNVEKQ 526

```


Oy	525	THHDPASJENSUKNLEKRLKALKWEDSRLNIYVEKELKYXKNLSIKENIEITLVEN	584
Db	527	RYNKFFSSNNSVNYOKLKKALSKYLEDSYLRGISEKQFNHYTLKGLLEDKIDKILTEE	586
Oy	585	IKKDEOQLFEKK---IIRKDNKSPDEKILEVSDIVKVOVOKVLMNKIDELKKTQOLIKNV	641
Db	587	IKKSENNKILEKFNFGILHSHANAS----LEVSDIVKIQOVOKVLLIKIDELKRIELFLKNA	642
Oy	642	ELKNNIHVPNSYKQENKQEBPYLYLVKLEKIDKLVMPVESLINEEKNKITEGOSDNS	701
Db	643	OLKSHIHVPNYKQPNKPEPYLYLVKLEKIDKLEKIPKRYDKMLKKEQAVLSS-----	695
Oy	702	EPSTEGELTGATTKPEGOAGSALDEDSVOAQOEOBKQ-----QPPYPRVPEAKQ	754
Db	696	--TIQPLVASETTEDEGSHSTHTL--SQSGETELETETEYEEVGHPTTVITITLPEKES	752
Oy	755	VPTPAPAVNKNTEN-----VSKIDLYLEKYLEFNTSYICHKYIIVASHSTPMNEKILQ	806
Db	753	APKEVYKVENSIIEKHSNDNSQALTIVYLVKLEKIDELFKTSYICHKIITIVSNSMDKILEV	812
Oy	807	YKTKKEBSKILSCDPLDLLFNIONNI PWAYSMFDSLNSLSQLFMEIYERKEMVCNLYKL	866
Db	813	YNLTPREENELKSCDPLDLLFNIONNI PWAYSLYDMDNIDLHFEELYOKEMAIYYLHKL	872
Oy	867	KDNOKINLLEAKV-----STYKTLSSSS-----MQPLSLTPQKPEVS	908
Db	873	KEENHKKLLEBOKQILGTGSTSTSPGGTIVYVNAQSAYTHSNSNOOSMASSTTQOGVANS	932
Oy	909	AND---DTSHS---TNLNSLSKLFENILSIGKKNKIYOEL--IGOKSENFEXELTKSD	960
Db	933	SGPAVVEESHDPPLTVLISNDKGIYSVLNKGKTKYVPRPLITSTTEMKFTEFLNKND	992
Oy	961	TFVNESTFNPYKSKADDINSLNDESKRKKLEEDINKLKTQOLSFDLYNKYKLLERLFD	1020
Db	993	TYFNDDIKQEFKYSMSKYITGLT--ETQKNALNDEIKKLDQLOSPDYKRYKXKLDRLFN	1051
Oy	1021	KKKRVGKKNQIKKLTLLKLOLESKUNSLNPNKHVLONSVFPNKKKAEIATENTLEN	1088
Db	1052	KKKELGODKQIKKLTLLKLOLESKUNSLNPNKHVLONSVFPNKKKAEIATENTLEN	1111
Oy	1081	TKILKHVKGIVKYVNGESSPLTKLSEESIOTEDNVAASENFKVLSKLEGLKLDQMLNEK	1140
Db	1112	TKILKHVKGIVKYVNGESSPLTKLSEVSIQTPEDNVAANEKRLASKIDGKLDNDMLHKG	1177
Oy	1141	KKLSYLSGGLHLLIAELKLEVINKKNYTGNSPSENNTDVNNALSKYKPLPECTDVAIVVS	1200
Db	1172	KKLSFLSSGLHLLITTELKEVYKKNNTGNSPBNKKVNEALKSYENFLPE--AKTVIYVT	1230
Oy	1201	E-----SSDPLLEOSQPKRKPASTHVGAESNTTITTSQNDVEDVDVITVIF	1246
Db	1231	PPQDVTYPPSLVARSVSSSGSTKEETQOIPSGLSLTELQOVOLVONEEDSLVPLPIF	1290
Oy	1247	GESEEDYDDLGQVVTGAVPSPVIDNLSKTEBEYEVULKPLAGYRSKIKOLENNWTF	1306
Db	1291	GESEDNDDEYLDQVYTGAIIVT--MDNLSGFENEEDYITIKPLAGYRSKIKOLEKNIT	1349
Oy	1307	FNVNVKDLNSRFNKRNFKNVLESOLIPYKDLTSSNYVAKPYFLNLEKKRDKFLSYN	1366
Db	1350	FNLMNLNDILNSRLKRRKYFLVDVLESQDMQFKHISNEVIIEEDSFILNSSEQNTLKSQY	1409
Oy	1367	YIKRSIDTDINFANDVIGYIKILSEKYSKSDLSIKKYT-----ND	1406
Db	1410	YIKSESVENDIKFAQOEGISYEVKVALKDDLESIKVILEEKKEPPSPPTPPSPAKTD	1469
Oy	1407	KOGENEKUYLPFLNNIEFLYTVVNDKIDLPVHLEAKVLWYTERKSNAVEKIKELMYLKTI	1466
Db	1470	EOKKESKFLPEPLNIEFLVNLNVKNIDDTVLINLKAICANQNEKQBAHKYTKLSIDLKAI	1529
Oy	1467	QDKLADRRKNNNFVGIADLSTDVNNHNNLLTKFLSEITGMVNEENLAKTVLSLNDLGNQGMN	1526
Db	1530	DDKIDFLKRNNDPEAIKKLINDTKKQDMIGKILSTGLV--QNPNNIISIKLIEGRKQDMYN	1588
Oy	1537	ISOHCYAKKQCPONSGCFRHLDERECCKLINTKQEGDKCVENPMPNCTENNNGCCADAK	1566

```

Db      1589  ISOHCYVKQCEHNSGCFRRHLDERECECKCLNAYKQEGKCYENPPTCNENNGCCDADAT 1648
|||||
QY      1587  CTEEDSGSGNGKRTCECTKPDSPYPLFDGTFCCSSNFIQSLFLILMLILYSFI 1639
|||||
Db      1649  CTEEDSGSGSRKRTCECTKRPDSYPLFDGTFCCSSNFIQSLFLILMLILYSFI 1701
|||||

RESULT  7
A26868
Major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C.Species: Plasmodium falciparum
C.Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jun-2000
C.Accession: A26868
R.Tanabe, K.; MacKay, M.; Goman, M.; Scaife, J.G.
J. Mol. Biol. 195, 273-287, 1987
A.Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmod
A.Reference number: A26868; MUID:88011243
A.Molecule type: DNA
A.Residues: 1-1701 <TAN>
C.Superfamily: major merozoite surface antigen
C.Keywords: surface antigen
F.1-19/Domain: signal sequence #status predicted <SIG>
F.20-1701/Product: major merozoite surface antigen #status predicted <MAT>

Query Match      60.0%; Score 5053; DB 2; Length 1701;
Best Local Similarity 60.1%; Pred. No.1.6e-149;
Matches 1042; Conservative 225; Mismatches 338; Indels 130; Gaps 30;

QY      1      MKIIFFLCSFLEFIINTQCVTHESYQELVKKLEALDVALVTGYSLFOKEMVNEGTSGT 60
Db      1      MKIIFFLCSFLEFIINTQCVTHESYQELVKKLEALDVALVTGYSLFOKEMVNEGTSGT 60
QY      61      AVTTSTPGSGSVASGSGGSAAGSGSAAGSGSAAGSGSGNSRRTNPSDSSDS 120
Db      61      AVTTSTPGSGSVT---SGGSAVASVASVSG--SGGSAAGSGSGNSRRTNPSDSSDS 114
QY      121      DAKSYADLKHRYRNLVTITKELVLPQLEFDLNNHMLTCDNTHGFEVLYLDGVEEINELLYK 180
Db      115      NTKYTAADLKHRYQNLVFTIKELVLPQLFDLNNHMLTSLKNVDGFKYLDGVEEINELLYK 174
QY      181      LNFYFDLLRAKINDVCANDYCOIPENLKRANELDVLKLVFGYRKPDLNIDKNVGKMD 240
Db      175      LNFYFDLLRAKLNDACANSYCOIPENLKRANELDVLKLVFGYRKPDLNIDKNVGKMD 234
QY      241      YTKKKKKKTENINELIESKKTIDKNKNATATEEEKKKIYQAOYDLSTYNNKOLEAHNLIS 300
Db      235      YTKKKKTTIANINELIESKKTIDONKNADNEGKKIYQAOYMLFTYNNKOLEAHNLIS 294
QY      301      VLEKRIIDLTKKNEIKELDLIKNEIK---NPPRANSGNPTNLLD-KNKKIEEHEKEIK 355
Db      295      VLEKRIIDLTKKNEIKELDLIKTDAENP---TTSKKNPLPENKKKVEEGHEKEIK 351
QY      356      ETAKTIKENIDSLFTDPLELEYLYREKKNKIDISAKVE--TKESTEPN-EYPNGVTPLS 412
Db      352      ETAKTIKENIDSLFTDPLELEYLYREKKNKVDVAPKSDPPKSYQVPIPVAPNGVPLP 411
QY      413      YNDINNAL---NELNSFDGLNPFDTYEPKSKNIYTDN-EKKKTIENEIKETIEKKKI- 467
Db      412      LTDINNSLAADNDKKSYSDDLMP--DTKEKINEKIITDKNKEKKIIFINNIKKOIDEERNIN 470
QY      468      ---ESDCKSYEDRSKSLNDITKEYEKLIENLYDSGNFNINIDLTNPEKKMGKRYSYATVK 524
Db      471      HTKEONKKLLDEYK---KKDYELLEKTYEMKFNPNPDVAVDIFESKARYTYNVEKQ 526
QY      525      THHNPFASYENSKHLEKLTALAKYMEDYSLRNIVVEKELKYUNLISKIENIETETLVEN 584
Db      527      RYNNKFNSSSNVYVQIKLTKALSLYEDYSLRKGISGEKDFNHYTYLTKGLEADIKKLTVE 586
QY      585      IKKDEQLEPKK---ITYDENKPRDEKILLESVDIYKVOYQVLLNKKIDELKLTOLLILKNV 641
|||||

```

```

Db 587 IKSEENKLEKFKGLTHSANAS----LEVSDIVKQVAVLLIKKIEDLKRIEELFKNA 642
Qy 642 ELKNIHVNSYKQENCOEYVLLVKKEDLKVFPMKVESLNEEKNKTKTEGQSDNS 701
Db 643 QLKDSIHVPNPKQKREPYLLVKKEDLKVEIPKVKMDLKEEDAVLSS-----695
Qy 702 EPSEGETGATTKPGQAGSALEGDSVQAQAEOKA-----QPPVAPVPEAKQ 754
Db 696 --IIOPLVASETTEDGSHHTL--SOSGEFEVIEFFVEETEGHTTIVTITLPRKEES 752
Qy 755 VPTPAPVNNKTEN-----VSKDYLEKLYEPLNTSYCHKYIIVSHSTNMEKILKQ 806
Db 753 APKEVKYVENSIEHKSNDNSQALTKTYVKLDEFLTKSYICHYIIVSNSMDQKLEEV 812
Qy 807 YKIKREESKLSGCDPLDLFNIONNIPVMSMEDSLNNSLSQFMEIYEKEMCNLYKL 866
Db 813 YNLPRKEEKLEKSCDPLDLFNIONNIPAMYSLYDSMNNDLOHLFEELYOKEMYYLHKL 872
Qy 867 KDNDRIKKLEAKKRV-----STSVKTLSSSS-----MOPSLTPQDKPEVS 908
Db 873 KEENHIKKLEEQKOITGSTSSSPGNTVTNTAQSATHSNQNOSNASSSTNTQNGAVAS 932
Qy 909 AND---DTSHS---TNLNSLKLEPNILSGKNKNYOEI--IQKSSSENYEKILKSD 960
Db 933 SGPAVVEESHDPVLTVLSINDLKGIIVSLDNGKTKVPNPLTISTTEMEKEFENILKND 992
Qy 961 TFVASEFTNFKSKADDNLSNDESKRKKLEEDINKLKTOLSPDYVNNKTKLERFD 1020
Db 993 TYFMDTIKQFKXSSSKYITGLT--EQKNALNDEIKKLDTLOLSDLNKKKLDRLFN 1051
Qy 1021 KKKTVGYKMOIKKLTLLKEOLESKLNSLNPNKHYLVNFSYFNNKKKAEIETENTLEN 1080
Db 1052 KKKELGQDMQIKKLTLLKEOLESKLNSLNPNHYLVNFSYFNNKKKAEIETENTLEN 1111
Qy 1081 TKILLKHYGKLVKYYNNESSPLKTLSESIOTEDNVASLENFKVLSKEGLKDNLEK 1140
Db 1112 TKILLKHYGKLVKYYNNESSPLKTLSEVSIOTEDNVAMLEKFRALSKIDGLNLDNLHLGK 1171
Qy 1141 KKLVSLSGLHLLAELEKVKNNKNTGNSPENNTPVNNLSESKKRLPGTQVATVVS 1200
Db 1172 KKLVSLSGLHLLAELEKVKNNKNTGNSPENNTPVNNLSESKKRLPGTQVATVVS 1230
Qy 1201 E-----SSDYLEGQOPKKPASTHVAESNTITTSQNVDEVYIIVPIF 1246
Db 1231 PPOQDVPSPSLVAVSGSGSGSTKEETQIPTSGSLTLEQVQVQONDEVDEDSLVLPF 1290
Qy 1247 GSESEDVDDLGQVYTGAVTSPVINDNLISKIENEYEVLYKPLAGVYRSLEKQLENNMT 1306
Db 1291 GSESDNDEYLDQVYTGAVTSPVINDNLISGFENEYDVLYKPLAGVYRSLEKQLENNMT 1349
Qy 1307 FNVVAKDILNSRFNKRRENFKNVLESDDLIPYKDLTSSNVVADPYKFLKKEKRDFFLSSYN 1366
Db 1350 FNLNLNLDLNSRLKRRKRYFLDVLDESOLMOPKHISNEYIIDSFKRLNSBOKNILLKSYK 1409
Qy 1367 YIKDSIDTINFANDVLYGYKYLSEKYSKSLDSIKKYI-----ND 1406
Db 1410 YIKESVENDIKFAEGEISYSEKVLAKYKXDESIKVIKEKEKFPSSPPTPPSPAKTD 1469
Qy 1407 KQGENEKYLPPLNNIETLYKLVNDKIDLEFVHLFAKVLNTYEKSNEVEKELNLYKTI 1466
Db 1470 EOKRESFPLPLTETLYLNNLVNKKIDYLLNLKAKINDCNEVEDEAVHKTJTKLSLDAI 1529
Qy 1467 ODKLADRRKNNPFGVADLSTDYNNHNLTLFTSGMY--PENLAKTYVLSLLDGNLQGM 1524
Db 1530 DDKLIDLEKNNIDFALIKKLMDTJKKMDLGLSTGLVQIIPN---TITSLLISGKFDOM 1586
Qy 1525 LNISQHCQVKKQCPQNSCFHHDREBCKCLLNYKQGDGKCVENPNTCENNNGCCAD 1584
Db 1587 LNISQHCQVKKQCPQNSCFHHDREBCKCLLNYKQGDGKCVENPNTCENNNGCCAD 1646
Qy 1585 AKCTEEDSGSGKAKITECTKPDSTYPLFDGIFGSSSNFLGISFLILLMLLYSTI 1639
Db 1647 ATCTEEDSGSGSKAKITECTKPDSTYPLFDGIFGSSSNFLGISFLILLMLLYSTI 1701

```

```

RESULT 8
A45604
major blood-stage surface antigen Pv200 - Plasmodium vivax.
C:Species: Plasmodium vivax
C:Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 24-Nov-1999
C:Accession: A45604
R:Gibson, H.L.; Tucker, J.E.; Kaslow, D.C.; Kretzli, A.U.; Collins, W.E.; Kieffer, M.C.
Mol. Biochem. Parasitol. 50, 325-333, 1992
A:Title: Structure and expression of the gene for Pv200, a major blood-stage surface
A:Reference number: A45604; M01D:92158013
A:Accession: A45604
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1751 <GIB>
A:Cross-reference: GB:M75674; NID:g160608; PID:g457336
A:Note: sequence extracted from NCBI backbone (NCBIN:83591, NCBI:P:83592)
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query: Match 36.7%; Score 3093.5; DB 2; Length 1751;
Best Local Similarity 38.2%; Pred. No. 7.6e-89;
Matches 704; Conservative 316; Mismatches 517; Indels 307; Gaps 41;

Qy 1 MKIIFLCSFLFIINQCVTHESYQELVKLELDAVLGYSLFQKVKVNLGTSQT 60
Db 1 MKALLPESLPEFYATKQCEET--ESYKOLVAKDLKLEALVVDGYELFPHKKL-----50

Qy 61 AVTSTFGSKGSVASGSGSVASGGSVASGGSVASGSGSGNSRTPNPSDSSDS 120
Db 51 -----GENDLKVEFNASANNNN 68

Qy 121 DAKSYADLKHRVNYLYLTIKELTPQLEPDLTNHMLTLCDNHIGKRYLIDGEEINELLYK 180
Db 69 NOVSV--LTSKIRNFKSLELEQIPGHTDLHLRELAVENPGIKYLVESYEENQLMHV 126

Qy 181 LNPFEDLRAKLVNVCANDQOIPFNLIKIRANELDVLKLVFGVRKPLDNTKDNVGMED 240
Db 127 INFHTDLRAKLVNVCANDQOIPFNLIKIRANELDVLKLVFGVRKPLDNTKDNVGMED 186

Qy 241 YIKRKKTIENINEL--TEESK-----TIDKNNATKEEKKKLYQAOYDLSYNNQ 261
Db 187 FYFNKTKITIKNIISLLIAENKKRSGHPTTNGAGTQPANGSIAASSETTQISGSSNSG 246

Qy 262 -----TIDKNNATKEEKKKLYQAOYDLSYNNQ 291
Db 247 SSGTSSNGSSSTSGSGTGTGQSPPAADASTNANYEAKKIYQAVYNTIFYTNQ 306

Qy 292 LEEAHNLISVLEKRIIDLKKNENIKELLDKINEIKNPPA---NSGNTPTLLDKNKKIE 348
Db 307 LOEAKLIVALEKRYKYLKEKDKLVLEQVAKKEKLPSPYPTNLTNVHKAESKIA 366

Qy 349 EHEKEIKETIANTIKFNIDSLFTDPLELEYLYREKNNKI--DISAKVETKESTEPNE----402
Db 367 ELEKKIEIAKTAVNFDLGLFTDAELELYLREKAKMAGTILIPESKRSAGTPEKTYPTL 426

Qy 403 ---YPNGVYPLSTNDINNMLNELNS---FEDLINPDPYKPEPSKNI--YDNEKAKKINE 455
Db 427 KETYPHGISTYALANSIYELEKIGSDTEFEDLQNP--DDGQOPKKGILINETKREKLEK 485

Qy 456 IKKIKIEKKIEESDKSYEDRSKSLMDIRKEYEKLNETIYDSKFENNIDLTNEKMMKG 515
Db 486 IMNKIKIEEDLPLMKYEIEKYVYEAKVNEFPANHFYEALDLNLTVENKFTDPRKK 545

Qy 516 RYSYKVEKLTHTNHFASYSKSHNLKLTAKLYMEDYSLENIYVEKELYYKYLKISKE 575
Db 546 REAVMEKKKIES--CSEYQNSNLIINKLKQLYTLEQVYLRKDLADEIKHFSMEKKLK 603

Qy 576 NEITLVENTIKKDEQOLFEEKITIDENKPDKEILVSDIVKVOYKVLNKKIDELKTKQ 635
Db 604 SEIYDLAOEIRKNNKLT-----TIENK-----PDFSGVVELQVAVLLIKKIEALKNVQ 652

```

0y	636	LILKNVLELNHNHVPNSYKOEKNQOEYYIIVLKELEIDKLKMPKVESL,INEEK-----	690
Db	653	NLKNMAYKDDDDLLPYKVKYKTEKPEPEYYILMVILKREIDKLKQEPKIESMIA,TERKNKPTVA	712
0y	691	--NKTGEGOS--DNSESTG-----EITQATTKRG-----QQGSA-LE	726
Db	713	AADIVAGGSLRGASLSEGTGTGNTVNAQIVAVOPROHOVANAVVQGGTGHQAQSGEAE,TO	772
0y	727	GDSVOAQAOEOKOA-----QPVPVPU-PE-AKAOVPTPAPVNNKNTENWS	771
Db	773	TNSVQAQAOVQOTFAGAGGQVASTQOTISQAPAPVQASPEAPAPASTMAA,VAAPRMSK	832
0y	772	LDVLEKYLEFLNTSYCHKIYLVSHSTNNEKILKOYKTKTEESKL--SCDPLDLLFNI	829
Db	833	LEYLEKLEIDLELSAYACHKHIFVNTSMKKEILDYKILNADQONKINTEKODELD,FTNV	892
0y	830	QNNIPVWYSMPDLSNLSLSQLEFMEIYERKMNONIKKLNDQ-KIKNILLEAKK-----VS	883
Db	893	QNNIPAMYSTYDSMSNLOLWLYELTQKDMYVNIYKKNQDKQKIAFLETKSKAAAPAQ	952
0y	884	TSVKTLSSSMOPLSILT-PODKPEVSANDTSHTNLSNLSLFEFNITSLCKNNKIYOEL	942
Db	953	SAAPRSQACTTYTTATPATTYTTVTPSPQTSVYSTPTTPPAEEN-RVAGNSEPEKE-	1010
0y	943	IGQSSSENFYEKILKDSDFYNESTFNEVFKSADDINSINDESKRKLEEDINKLK,TLQ	1002
Db	1011	ADTAQVKEPEKHELSQIDK-YNDYFQKPFESQDEITTK-DETKKALGALFEELKKI,Q	1068
0y	1003	LSFPLKYKYLKLELRFDKKKTKYTKKMOIKTKLTLKDLQESKLSLNLPKHVLO,NSVF	1062
Db	1069	VSDHYKAYKYLKEERLLKKNKINSKDDIKRLTSLKNLEKROWLNNPVS,LYKNTAF	1128
0y	1063	FNKKKEAIEETENTLENTFNTILKHKYGLVYKYNQESSPLKTLSESIOTQD,NA,SENF	1122
Db	1129	FNKKRETEKEVENTLKNLEIKLKYKAKAKYIIGEPPLKTLSESKQKEDNTY,LEKF	1188
0y	1123	KVLSLEGLKLDNLENEKKKLSYLSGSLHHLIAELKEVYKNNKNTYGNSPSEN,NTDVNAL	1182
Db	1189	RVLSELRGRGKNIETELEKENISYLSGSLHVLTELKEIKNNKYSGNDHTNIAA,VKRAL	1248
0y	1183	ESYKFLPEGTDAVYVSSGDTLEQS-----QPKPA-	1216
Db	1249	QAVOELPRK-----VTQEGASTTAA,TPVTPVPSAVPGLFAGAVPAAAGL,PPAPAG	1302
0y	1217	---STHGA-----ESNTITTSQNVQDEVDVITVPIRESEEDVD,IGOVYVGA-----	1264
Db	1303	SVAPTEGGAAGSTEENVAAKADVAEDVDKYALPREFNNDDEE--DOYTTGSE,SEA	1361
0y	1265	---TTPSEVIDNII,SLKENEXEYVLEPLAGVVRSLKQJENNVMFNVNVDI,INSRENK	1321
Db	1362	PELIVPAGI-----SDDDVYILKPLAGMKKTKIKQJENNVNMFNNTIDM,DSRLK	1413
0y	1322	RENFKNVLESDLIPYKDLJSSNVKQPYKFLNKEKRDKEFSSYNYIKDSIDT,DI,INFAND	1381
Db	1414	RNYLEVLANDLMPFYKSSSGEYIITKDPKLLDEKKKLLIGSYKYIGASIDM,DLATAND	1473
0y	1382	VLGYYKTLSEKYS,SDSIK--KYIND-----KQEN	1411
Db	1474	GVTYVYNNMGEVLTHYHDGVYTEIKVVEDIKQDELEKLGKGNVSQDSKKNEFIAK,AEI	1533
0y	1412	EKYLPRFLNNTIETLYKTVNNDIDFLVHLBEAKV,LYNTEKSNYEVKIKRELANT,KT,IODKA	1471
Db	1534	EKYLPRFLNSLOKREYESLVSVNYYTONLKKVNNINOLEKKEKEITVYKLODYKNK,DELE	1593
0y	1472	DFKKNNNFVGID,LSIDYVNNHNLKFLSTGWFENLAKTV,LSNLDGNLOGM,NISQHQ	1531
Db	1594	EYKK-----SEKKNVKSSGLLEKLMKSKLIKENESKEIISOLLANTO,LT,MTMSSEHT	1646
0y	1532	CYKQKQCPQNSGCFRPHIDERECKCLINTYQOEDKCYEVPNPTCNENNGCQADAK,CTEED	1591
Db	1647	CITNVVDNACRYKIDGTEEMKCLLTFEYEBQKCPVPSAVNYCKQNNNGCA,PEAEKWTJD	1706

[illegible]


```

QY 1192 GTDAATV-----SESGDTLEOSOPKPASTHV-- 1220
| | | | |
Db 1226 STSAVAVVPPAGVPTAAAGSGAGAVPPAAAGSGAGAVPPAGGPPPTGCVVP 1265
| | | | |
QY 1221 ---GASNTITTSQNVDEVDVITVIFGESEEDYDGLQVVTGEA-----VTPSV 1269
| | | | |
Db 1286 GVSASAEQKFAQADYAEVDYIELPLFGNDDGEE-DQVTTGEASEAPLIVPAG 1344
| | | | |
QY 1270 IDNLSKTEENYELLYPLAGVYRSLSKQLENNVMTFNNAVKDIILSRFKENKAVL 1329
| | | | |
Db 1345 I-----SDYDVVYLKPLAGMYKIKQLENNVNAFNMTITMDLSRLKKNVFLVVL 1396
| | | | |
QY 1330 ESDLPKDTSSVYVVDKPKFLNKEKROKFLSYNYIKSIDTIDFNAVADVGYKIL 1389
| | | | |
Db 1397 NSDLNPKRPSGSEYITKDPYKLLDLEKKKKLSTYKIGASIDKGLTADGVNYITKM 1456
| | | | |
QY 1390 SEKYSDL-----DSIKKYIND-----KOGENEKYLPLN 1419
| | | | |
Db 1457 GELKTHLTVNEEVKVEADIKAEDEKIKKIGSDTKTTEKTSMAKKALEKYLPLN 1516
| | | | |
QY 1420 NIEPLYTVNDKIDLEVIHLEAKVLTYYEKSNEVYKIKELNYLKTODKLADFKKNNF 1479
| | | | |
Db 1517 SLOKEYSIVSKVNTYTDNLKKVINNCOLEKKEAEITVKKIJDYNNKDEKLEEYKK--- 1572
| | | | |
QY 1480 VGADLSTDVYNNHLTKFLSTGVFENLAKTVLSNLLDGLQGLMINSOQYKQKCPQ 1539
| | | | |
Db 1573 ---SEKKNVKSGLLEKLMKSLIKENESKEILLSOLLVOTQLTMSSEHTCIDTNP 1629
| | | | |
QY 1540 NSGCFRLDEHEECCKLNTYKQEGDKCENPNPCNENNGCGDADAKTEDESGSNGKI 1599
| | | | |
Db 1630 NAACYRILDGHEMKRCLITREBEGCKVPGSNVTCXKDNNGCAPAEKCKMDS-----NKI 1665
| | | | |
QY 1600 TCECTPDSYPLFDGIFCCSSNPLGISFLILMLIL 1635
| | | | |
Db 1686 VCKCTEGSEPLFEGVFCSSSSFLSLILMLLFL 1721
| | | | |

```

RESULT 11

506286
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
N:Alternate names: 190K protein; polymorphic schizont antigen
C:Species: Plasmodium falciparum
C>Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 09-Jun-2000
C:Accession: S06286
R:Certa, U.; Kotmann, D.; Matile, H.; Reber-Liske, R.
EMBO J. 6, 4137-4142, 1987
A:Title: A naturally occurring gene encoding the major surface antigen precursor p190 of
A:Reference number: S06286; MUID:8816657
A:Accession: S06286
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-1060 <CER>
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 33.6%; Score 2829.5; DB 2: Length 1060;
Best Local Similarity 55.5%; Pred. No. 6.5e-81;
Matches 609; Conservative 145; Mismatches 255; Indels 89; Gaps 22;

```

QY 1 MKIIFLCSPFLFIINQCVTHESYOELVKLELADAVLNGYSLFOEKWLVNGSGT 60
| | | | |
Db 1 MKIIFLCSPFLFIINQCVTHESYOELVKLELADAVLNGYSLFOEKWLVNGSGT 60
| | | | |
QY 61 AVTSTPGSKGSAVSGSGVSAAGSVASGSGVSAAGSGSGNRRTPNSDSSDS 120
| | | | |
Db 61 VVAPPADAVSTQSAKKNPGCATVPS-----GTASTKGAISPSG-----ANPSDDSSDS 108
| | | | |
QY 121 DAKSYADIKHVRNLTIKELKLPOLFDTLNNMLTLCDNHGRKYLLIDGYEELINELLK 180
| | | | |
Db 109 DAKSYADIKHVRNLTIKELKLPOLFDTLNNMLTLCDNHGRKYLLIDGYEELINELLK 168
| | | | |

```

```

QY 181 LNFYFDLLRAKLVANDYCOIPFNLIKIRANELDVLKLVFGYRKPLDNKIDNWKMED 240
| | | | |
Db 169 LNFYFDLLRAKLVANDYCOIPFNLIKIRANELDVLKLVFGYRKPLDNKIDNWKMED 228
| | | | |
QY 241 YIKRNNKTINELIEBSKKTIDKNKATKEEKKKLYAOYDLSYNNKOLEBAHLNIS 300
| | | | |
Db 229 YIKRNNKTIANINELIGSKSKTIDQNNKADNEBCKKLYAOYDLSYNNKOLEBAHLNIS 288
| | | | |
QY 301 VLEKRITPLKNNENIKELDLKINIKNPPRANSNGTNTLLDKKKKIEHEKELIKETAKT 360
| | | | |
Db 289 VLEKRITPLKNNENIKELDLKINIKNPPRANSNGTNTLLDKKKKIEHEKELIKETAKT 348
| | | | |
QY 361 IKFNIDSLFTDPLEYLRREKNKNIDISAKVE--TKSETPN-EYNGVYPLSYNDIN 417
| | | | |
Db 349 IKFNIDSLFTDPLEYLRREKNKNIDISAKVE--TKSETPN-EYNGVYPLSYNDIN 408
| | | | |
QY 418 NAL---NELNSFGDLINFDYTKEP-SKNITYDN-ERKKFINEIKELIKETAKT---E 468
| | | | |
Db 409 NSLAADNDKNSYGLMNP--HTKEKINEKITTDNKKERKIFINIKIKQIDLEKNINHTKE 466
| | | | |
QY 469 SDRKSYEDRSKSLNDITREYKELNIYDSKFNNDIDLTFEKKMMGRYSKYVEKLTHHN 528
| | | | |
Db 467 QNKRLEDEYERS---KDYELLEKEFEEMKFNNNFNKDVDYDKIFSARYYVVEKORYNN 522
| | | | |
QY 529 TFAVENSKNHLEKLTALKYMEDYSLRNIVEKELKYKNLISIKENIEFTLVENIKKD 588
| | | | |
Db 523 KFSSNNSVYNVOKLKALSTLEDYSLRKGISEKDFNHYITLKGLEDIKKLEETIKSS 582
| | | | |
QY 589 EEOLEFEK---ITRDENKPDKEILEVSDIVYVOYVOKVLLNNKIDELKKTQDILANVELKH 645
| | | | |
Db 583 ENKLEKNFKLTHSANS---LEVSDIVYKLOVQVLLIKIKIEDLRIELFLNNAOLKD 638
| | | | |
QY 646 NIHPVNSTKQENKOPYYLIYKKEIDKLVFMPKVESLIEEKKNNKTIEGOSDNSEPT 705
| | | | |
Db 639 SIHPVNIYKPKNKEPPYLLIYKKEIDKLVFMPKVESLIEEKKNNKTIEGOSDNSEPT 689
| | | | |
QY 706 EGEITGQATTKPGQAGSALGDSVQAOBOKO-----AOPVPVPEAKAOV 755
| | | | |
Db 690 QPLVAASSETTBEGGSHHTLSQSETEVTEETELVGHHTVTITLPPKEKAVYENSTEH 749
| | | | |
QY 756 PTPAPVNNKTENYSKLDYLEKLYEFLNTSYICHKYLIVSHSTNNEKILKOYKTKKEES 815
| | | | |
Db 750 KS-----NDNSQALTKYVLLKLEDFLTKSYICHKYLIVSSNDOKLVEYNYLT-PEEN 803
| | | | |
QY 816 KLSCDPLDLLFNQNNNPVWYSMPDSLNSLSQLFMEITKEKWCNLYKIKADKDKITNL 875
| | | | |
Db 804 ELKSCDRLDLLFNQNNNPVWYSMPDSLNSLSQLFMEITKEKWCNLYKIKADKDKITNL 863
| | | | |
QY 876 LEEAKV-----STSVKTLSSSS-----MQPLSLTPQDKREVSAND---DTS 914
| | | | |
Db 864 LEEPKQITGSTSSPGNTVTYNTQASATHSNQOQASASTNTNONGVAVSSGPAVVEES 923
| | | | |
QY 915 HS---TNLNSLSLLENILSLGKNKNIYOEL-IGQSSSENFYERKILKSDOTFYNESEFTN 969
| | | | |
Db 924 HDPLTVLISNDLKGIVSLMLNKNKTKVPNPLTISTEMEKRYENILKIMIPENDIKQ 983
| | | | |
QY 970 FVSKADINDSLNDESKKKKLEEDINKLKTQLSFDLYNNYKTLKLELFPKKTATVGYK 1029
| | | | |
Db 984 FVKSISKVITELT-ETQNNALNDELKIKLDTLQSLFYNNYKTLKLDLRFNKKKELGDK 1042
| | | | |
QY 1030 MQIKKTLTKLEQLESKLN 1047
| | | | |
Db 1043 MQIKKTLTKLEQLESKLN 1060
| | | | |

```

RESULT 12

A45532
major merozoite surface antigen precursor - Plasmodium yoelii
C:Species: Plasmodium yoelii
C>Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 26-Aug-1999
C:Accession: A45532; A45531
R:Lewis, A.P.
Mol. Biochem. Parasitol. 36, 271-282, 1989

Db 1693 KGEQNTCVENNPTCDJINGCDPTASCQNAESTENSKIICTCKEPTPNAYEGVCCSS 1752
Oy 1620 SNFGISFLLIMLILYS 1637
Db 1753 SPMGLSILLITLIVFN 1770

RESULT 13
A45546
major merozoite surface antigen precursor - Plasmodium chabaudi chabaudi
C:Species: Plasmodium chabaudi chabaudi
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 26-Aug-1999
C:Accession: A45546
R:Deleersnijder, W.; Hendrix, D.; Bendahman, N.; Hanegreets, J.; Brils, L.; Hamers-Caste
Mol. Biochem. Parasitol. 43, 231-244, 1990
A>Title: Molecular cloning and sequence analysis of the gene encoding the major merozoite
A:Reference number: A45546; M0ID:91218803
A:Accession: A45546
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1785
A:Cross-references: GB:M34947; NID:q160597; PID:q160598
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 29 8%; Score 2509; DB 2; Length 1785;
Best Local Similarity 31.6%; Pred. No. 9.8e-71;
Matches 598; Conservative 324; Mismatches 605; Indels 366; Gaps 35;

Oy 1 MKIIFLCSPFLFIINTQCVTHESYOELVKKLEALDVAVLGYSLFOKEKKVNEGTSGT 60
Db 1 MKKIGLFSFVFPAIYKSETIGYVNDLVHKLKLELSEVGLLEFQKSOYIVAAOSPET 60
Oy 61 AVTTSTPGSGVASGSGSGVASGSGVASGSGVASGSGSGNSRTNPSDSSDS 120
Db 61 PVDPF-----TNP----- 68
Oy 121 DAKSYADLKHVRNRYLTLEIKLPOLFDLTNHLTLCDNHGFKYLIDGEEELNELLYK 180
Db 69 -----EFAOKLDPFLIKFELGFTTEOTELVNLKLTGPNKYGLKYLESEKEEENELMHA 122
Oy 181 LNFYFDLLRAKLDVANCANDYCOIIPNLKIRANLDELVLKLVFGYRKLDPNLIKDNVGMED 240
Db 123 INFYDYDLROKLDNMCANNYCEIPEHLKINVEEIMLKVYGRKPIENIQDVLLEE 182
Oy 241 YIKKNKKTINELIEESKKT-----DKKNATPEEEKKLYQAOYDLSTYKQ 291
Db 183 YIARNKATAEFLNLTLEETFKITPEEETDCNDTNCNTYKGGKKAITYQAMYNVIFYKKQ 242
Oy 292 LEEAHNLISYLEKRIDPLKKKENIKELDKINIKNPPRANS----- 334
Db 243 LAELKVIYELERKVALTKKNEAIKPLLOQLEAIRPPATEGOIATEGSEETKONSTE 302
Oy 335 -NTPNTLD----- 342
Db 303 SSMTKTTTTKAVTQTQATKATGETNTGTETNGTETNATGTTATGTTATGTPVT 362
Oy 343 -----KKKKEEHEKEIKETAKTKEINIDSLFTPLLELYLRE-KKKNI 386
Db 363 EPVOVPAVOVLTEEEKKKAIAELYAOIKELAKTIKFNIDGIFVDPVLELYKKKKKKNESC 422
Oy 387 DISA-----KVEETKSTEP--MEYNGVTPYLS---YNDINNALNELMSFGDLIPPPYTK 437
Db 423 HSHSSCKKNTPEVPLPNRYRPGISYPLTEEVVYSKIAHNAE-ITTYGDLTN-VNFTA 480
Oy 438 EPSKNIYTDERRKFINLEIKIKIEKKKIESDKSYEDSKSLNDITKEYEKLELNEYD 497
Db 481 ITTEDLTNEGARKNLKAIKIKKIEAEQKLVLELDDYDTFLAARNGOKTPKEKAKKFYE 540
Oy 498 SKFRNNIDLTNFEKAMGKRISYVEKLTHTNTRASYENSKHNEKTLKALKYMEDYSLRN 557
Db 541 SKFRNKLTTIDFDKFKRTREYNNKAA--LVGCEYNTQOOLINKLKOJNYLQDYGLRK 598

Oy 558 IYVERELKYKNIISKIENETETLVENIKKDEQLEFKKITKDKENKDEKILEVSDIYKV 617
Db 599 EIVNTEIEFESNKKSELOYNINRLANVAQAKONILVA-----SKHIFSTLVEL 647
Oy 618 QVQFVLLMINKIDELKKTQOLILKANVELKHNIHVPNSYQKQENQBPYVLLVLEKKEIDKLKVF 677
Db 648 QIQKSLTLKLEQLNTEKTEFSLNKAHLKDKIYVPQTYGEGKPEPYVILAIKKEIDRLAKE 707
Oy 678 MPKVESLINEKKNIKIEG-QSDNSPSTSEETITGQATTKGQQA----- 721
Db 708 IPKIDIMEKQKMEQEBHVATGSEQAASSASGSGSTETTSQTPAPVAPAPAEKAKE 767
Oy 722 -----QSALEGDSV---QAAQOKOQAP-----PVPVPEAKAQVPT 757
Db 768 GTESTEETPAASKAPAEAGASTGATTPTEQEAAPTEQEAQAPAPETPAEVPAPPTPAAPAT 827
Oy 758 PPAFVNKNKTENVSKLDYLEKLEFELANTSYICHKYLIVSHSTNNEKILKOYKITREES-- 815
Db 828 PAAPAAPAKPMTKLYLEKTKFLAFSYACHKVVLLQNSTINIKDALSKVALTPREBKIR 887
Oy 816 KLSSCDPLDLFNQNNIPVWISMFDSLNSLSQLEMEIYEKEVNCILYKLD-NDKIKN 874
Db 888 TLKRCSELDVLAIQNNMPTWISLENVVDLQNIYTELKEKMMHYHYNLKDKNPAVKA 947
Oy 875 LL-----EAKKVSTSVKT- 888
Db 948 LLYVAGYIDPEPPVAPPTPAPEAPETAPETAPETAQEAQAPQESQAQEAQETETTP 1007
Oy 889 LSSSMOPLSLTPQDKP-----EVSANDTSH 915
Db 1008 AESASTETPTAPATPATPTSERVTOEGTTPAPARKQEGASSAPAPAPAKAPATQITCQ 1067
Oy 916 STNINSLKLEFNILSLGKNKNITYOELIGOKSSENFYEKLIKDSOTYENESFTNPFVSKA 975
Db 1068 STNVEGSTQV-----RASEDEMEVDPEVDNPFKSYLQOVDG-NNQOIFDIFIKSK 1118
Oy 976 DDINSLDESRRKLEEDINKLTKTLOLSPDLYNKKYKLEKLELDFDKKTKYKYMOKIKL 1035
Db 1119 ELINALPPE-KVNDLYDIDIAHLKELSEHYIKRYKIKYKLELRLQKHQOILEANOKYKEI 1177
Oy 1036 TLLEQLESKLSLNSLNPKHVLONFSPFNKKKEAEIETENTLENTLNTILKHYGLVYKY 1095
Db 1178 SVLSRLKLRKKYVIGFVYLSGFANFPNKRREAEKQYVONAIKNTDMLKYYKARSYF 1237
Oy 1096 NGESSPLKTLSEESIOTEDNAYSLNFKVLSLEGLKLNINLEKKLSTYSSGLHLHIA 1155
Db 1238 TSEAVPLKTLTKTSIDREANLYKIEKFRAYSRLRLKLNINLGEKERTIYVSGGLHHVFE 1297
Oy 1156 ELKEVINKNNTYTGNSPSENNTDVNNALESYKFLPEGFD-----VATVYSSGSDTLEQS 1210
Db 1298 EFKELLNKGTGTGNTENPENAPEVIAKAEQYKELLPKGATTPAPVAVVAPAPATAAPA 1357
Oy 1211 QPKRP-----ASTHGAESNTITTSQNVDDVDVYIIVPIFGSEEDDYDLQO 1258
Db 1358 DAPVPAAPAAAAAGSGSAAATGEGEATTVVASSDND-----DDDDMDQ 1402
Oy 1259 VYTGAEATPVSYINDILSKIENEYEVLYKLPLAGYRSLKOLENNWTFENVNVVDILNSR 1318
Db 1403 IANAOQ--TDERVKDILDAFSENEIYITKSLGNTYKSEFKKIMLEFESKIKEDIMTGILNYK 1461
Oy 1319 FNRKRNKFNKLVESLDLIPYKDTSSNYVVKDPKFLNKKRPFSSYNYITDSIDTQINF 1378
Db 1462 LEKRNDELIVYSTELAFKQINTNKKFVYKPNQYLLDDDKDKQOMINLKYAIKGTEDLET 1521
Oy 1379 ANDVLAGYKTLSEKYSDDLDSIKRYI-----NDKQGENEYKLPPLANNIETLYKTVNDKID 1433
Db 1522 ATDGIIEFFNKMIELYKPOLNAVNEQIAIGTEPTDAKKKAPAFEDLKGLEYITILNGAE 1581
Oy 1434 LFVHLKAKVLYNTYKESNNEVYKIKELNYLKTJODKLADF-----KKNNFVGIADLSDY 1489
Db 1582 EFSLELOHKLENYKIEKAGDIIIMANLEYIYRIDEKLEDFVESAEKKNHIAISIA----- 1635

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2001, 13:14:34 : Search time 23.06 seconds
(without alignments)
1463.468 Million cell updates/sec

Title: US-09-269-874-3
Perfect score: 8424
Sequence: 1 MKRIFFLCFLFFIINTQCV.....SNFLGISFLILMLILYSFI 1639

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query length	ID	Description
1	615	7.3	115	US-08-152-922A-7	Sequence 7, Appli
2	356.5	4.2	3248	US-08-353-700-1	Sequence 1, Appli
3	356.5	4.2	3248	PCT-US93-16216-1	Sequence 1, Appli
4	351	4.2	1786	US-08-973-462-8	Sequence 8, Appli
5	350.5	4.2	2482	US-08-328-254-6	Sequence 6, Appli
6	332	3.9	976	US-09-104-324B-4	Sequence 4, Appli
7	308	3.7	1312	US-08-687-080-51	Sequence 51, Appli
8	307	3.6	1312	US-08-592-126-148	Sequence 148, App
9	301	3.6	53	US-08-290-919-3	Sequence 3, Appli
10	298.5	3.5	2710	US-08-405-496A-6	Sequence 6, Appli
11	298.5	3.5	2710	US-08-405-496A-6	Sequence 6, Appli
12	298	3.5	3135	US-08-323-170B-2	Sequence 2, Appli
13	296	3.5	1588	PCT-US93-07261-11	Sequence 11, Appli
14	296	3.5	1663	PCT-US93-07261-16	Sequence 16, Appli
15	293.5	3.5	2391	US-08-446-855A-2	Sequence 2, Appli
16	293.5	3.5	2391	US-09-150-741-2	Sequence 12, Appli
17	292.5	3.5	3418	US-08-755-587-44	Sequence 44, Appli
18	288.5	3.4	3418	US-08-603-753D-4	Sequence 4, Appli
19	288.5	3.4	3418	US-09-089-753-4	Sequence 4, Appli
20	288.5	3.4	3418	US-08-986-106-4	Sequence 4, Appli
21	282	3.3	53	US-08-290-919-4	Sequence 4, Appli
22	280	3.3	48	US-08-290-919-12	Sequence 12, Appli
23	279.5	3.3	1589	US-09-356-952-4	Sequence 4, Appli
24	279.5	3.3	2329	US-08-755-587-16	Sequence 16, Appli
25	279.5	3.3	3418	US-08-639-501-2	Sequence 2, Appli
26	279.5	3.3	3418	US-09-044-946-2	Sequence 2, Appli
27	279.5	3.3	3418	US-09-044-908-2	Sequence 2, Appli

28	278.5	3.3	2366	US-08-480-604A-10	Sequence 10, Appli
29	278.5	3.3	2366	US-08-405-496A-10	Sequence 10, Appli
30	275.5	3.3	1093	US-09-315-793-52	Sequence 52, Appli
31	274	3.3	48	US-08-290-919-1	Sequence 1, Appli
32	274	3.3	1354	US-08-685-871-2	Sequence 2, Appli
33	272	3.2	1535	US-08-755-587-185	Sequence 185, App
34	271	3.2	48	US-08-290-919-2	Sequence 2, Appli
35	271	3.2	2154	US-08-841-349-4	Sequence 4, Appli
36	270	3.2	2052	US-08-790-912-2	Sequence 2, Appli
37	269	3.2	1435	US-08-568-459A-4	Sequence 4, Appli
38	269	3.2	1435	US-08-487-826B-4	Sequence 4, Appli
39	267.5	3.2	1388	US-08-685-576-4	Sequence 4, Appli
40	266.5	3.2	1579	US-08-755-587-184	Sequence 184, App
41	265	3.1	1363	US-08-425-061-23	Sequence 23, Appli
42	265	3.1	1363	US-08-825-886-23	Sequence 23, Appli
43	265	3.1	1852	US-08-425-061-24	Sequence 24, Appli
44	265	3.1	1852	US-08-825-886-24	Sequence 24, Appli
45	265	3.1	1863	US-08-425-061-16	Sequence 16, Appli

ALIGNMENTS

```
RESULT 1
US-08-152-922A-7
: Sequence 7, Application US/08152922A
: Patent No. 5395614
:
: GENERAL INFORMATION:
: APPLICANT: Knapp, Bernhard
: APPLICANT: Knapp, Erika
: APPLICANT: Eenders, Burkhard
: APPLICANT: Kuemper, Hans
: TITLE OF INVENTION: Protective Plasmodium Falciparum
: TITLE OF INVENTION: Hybrid Proteins which Contain Part-Sequences of the Malaria
: TITLE OF INVENTION: Antigens HRP1 and SERP, the Preparation and Use Thereof
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
: ADDRESS: Dunner
: STREET: 1300 I Street, N.W. Suite 700
: CITY: Washington
: STATE: D.C.
: ZIP: 20005-3315
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/152,922A
: FILING DATE: 16-NOV-1993
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/806,471
: FILING DATE: 13-DEC-1991
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Binaudi, Carol P.
: REGISTRATION NUMBER: 32,220
: REFERENCE/DOCKET NUMBER: 02481-1143-02000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-408-4400
: TELEFAX: 202-408-4400
:
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 115 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
:
: US-08-152-922A-7
:
: Query Match 7.3%; Score 615; DB 1; Length 115;
```

Best Local Similarity 100.0%; Pred. No. 2.2e-26;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 KELVYPOLEDTLNMHLTCDNHGKRYLDGVEEINELLYKLNFPDLRLKNDVAND 199
Db 1 KELVYPOLEDTLNMHLTCDNHGKRYLDGVEEINELLYKLNFPDLRLKNDVAND 60
QY 200 YCOIPFNKIRANELDVLKLVFGYRKLPLNDKDVGMEDYIKKKKTININE 254
Db 61 YCOIPFNKIRANELDVLKLVFGYRKLPLNDKDVGMEDYIKKKKTININE 115

RESULT 2

US-08-353-700-1

Sequence 1, Application US/08353700

Patent No. 5599919

GENERAL INFORMATION:

APPLICANT: YEN, TIMOTHY J.

APPLICANT: RATNER, JEROME B.

TITLE OF INVENTION: NUCLEIC ACID ENCODING A

TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,

TITLE OF INVENTION: AND METHODS OF USE

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: DANN, DOEFMAN, HERRELL AND SKILLMAN

STREET: 1601 MARKET STREET, SUITE 720

CITY: PHILADELPHIA

STATE: PA

COUNTRY: USA

ZIP: 19103-2307

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/353,700

FILING DATE: 09-DEC-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: REED, JANET E.

REGISTRATION NUMBER: 36,252

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 563-4100

TELEFAX: (215) 563-4044

INFORMATION FOR SEQ ID #0: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3248 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHEICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: HUMAN

US-08-353-700-1

Query Match 4.28; Score 356.5; DB 1; Length 3248;
Best Local Similarity 18.94; Pred. No. 8.2e-11;
Matches 350; Conservative 304; Mismatches 631; Indels 569; Gaps 79;

QY 125 YADKKHR-VANNYLLITKELKYPQLFDLTNMLTCDNHGKRYLDGVEEINELLY 179
Db 709 FSDKHQKEIENMCKLTSQTL-GQVEDL-EHKLQLSN-----EIMDKRCYODLHAWE 761
QY 180 KLVNFPDLRLKNDVAND-----YCOIPFNKIRANELDVLKLVFGYRKLPLNDI 231
Db 762 SLR---DLKSKDASLYTNEHQSLLAFDQQRAMHSHFAN-----ITGEGSMSE 810
QY 232 KDNVGMEDYIKKKKTININEIESKKTIDKNKNATKEEKKLYQAOYDLSYVK 291

Db 811 RSECHLEADQSPKNSAILQNRVDSLEFLES-OKOMNSDLQKCEELVQIKGEIENLMK 869
QY 292 LEEAH-NLISVLEKRIIDLKKN-----ENIKELDKINELKNPPASGNTPTNLL 341
Db 870 AEOMHOSFVAETSORISKLQEDTSAHONVAETLSALNKKKEILO-----LL 916
QY 342 DKNKIEEKEIKEIAKTIFNIDS-----LFTDPLELEYLREKKNIDISAKVETKE 396
Db 917 --NKKVELEQAEIQELKSNHLLSDLSLEKLOLSETLSLE--KREMSIISLN-KREEE 971
QY 397 STEPNRYNGVTPPLSYDINDNALNELNSFDLINFEDYTPKSPKNITDNEKKKFINET 456
Db 972 LTOENG-----TLKEINASLNQEKM--NLIO-----KSESFANYID-EREKSISEL 1014
QY 457 KEKIRIEK-----KKIESKRSYEDRS--KSLNITKYEKELNFIYSKRNNNIDLTNE 509
Db 1015 SDQYKOEKLIILQCEETGNAYEDLSQYKKAQENSKLEGLNCSYLCENRKNLEQL 1074
QY 510 EKMCKRYSYRVEKLTJHNTFASYENSK--HNLEKLTALK--YMEDYSLRNIYVEKELKY 566
Db 1075 KEAFKHEQEFUTKL---AFAERNQMLMELETVQALSEMTDQNNKSKSEAGGLKQ 1130
QY 567 YKNLSKTIENIEPLVENIKKDEQLFEKKTITKDE-----NKPDEKILEYSDI-- 614
Db 1131 EIMTLKEBONKQKQEVNDLQENBQMLKVMKTKHCQNLSEPIRNSYKERESEBNOQNE 1190
QY 615 --YKVOYQKVLNKKIDELKKTOLILKNVELK-----HNHV----- 649
Db 1191 KPOMDLEYKEIISDSYNAQVLEAMLNKELKLOESKEKECLOHELOTTRGDLEISNL 1250
QY 650 -----PNSYKQENKQEPYLLIVLKEIKELDKLVMPKVESL 684
Db 1251 QDMQSOEISGLKCEIDAEKXYISGPHELSTQNDNAHLQSLQTTNKNLNE-LEKICEI 1309
QY 685 INEEKNKITTEGQSDNSPSTBGETGQATKPGQOAG-----SALEGBSVQ 731
Db 1310 LQAEKVELYELTBLNDSRSCIT-----ATRKMAEEVKKLNEVKILINDSGLLHGEI 1362
QY 732 ---AAQOEKQAOQPPVPPVPEAKQVTPPAPVNNKTENWSKLDYEKYEPLNYSYI 787
Db 1363 DIPGGEFEQNPNEQHPSVL-----APIDESNS-----YEHLLTIS-- 1386
QY 788 CHKYTLVSHSTMNKIL---KQYKTYKEESKLSLSCDPLDLLFNIONNIPWYSMPDSL- 843
Db 1397 -DKVEQMHFAELQEKFLSLQSEHKILHDHQOMSS-----KMSLQTYVDSLK 1443
QY 844 -NNSLSQLFMETYEKEMVCNLYKLKNDKYNLLEBAKVSTYKTYSSSSMQ----- 895
Db 1444 AENLYLSTNLNFGQDLVKEQMLGLEGLVPSLSSSCVPSDSSLSLSDSSFYRALLEQT 1503
QY 896 -PLSLTPQDKREVSAN-----DQTSHTNINNSLKLFINITSLG 933
Db 1504 GDMSLSLNLEGAVSANQCSVDEYFCSSLYQTYVDSIKAKENLYLSTNLNFGQDLVKEQMLG 1563
QY 934 KKNKIYDEL-----IGOKSENFYEKILK---DSDTFYVN-ESFTNIVKSKADDI- 978
Db 1564 LEEGLVPSLSSSCVPSDSSLSLSDSSFYRALLEQTMGMSLSNLEGVASANQCSVDEYF 1623
QY 979 -NSLNDSEKRRK-----LEEDINKLAKTLOLSPDLNKKYKLAERL 1018
Db 1624 CSSLOEENLTKREKPPSPAPKAVGELESLCEVYRSLKELEKMESSQGIKMKKEIOLEOL 1683
QY 1019 FD-----KKRTYGVKKMQIKKLTLLKQLESKLS-----LNKKHYLON 1058
Db 1684 LSSERQELCLRKOYLSNEMQOKLTYSVLEMSKLAEEKQTEQLSLEVARLDQ 1743
QY 1059 FSV-----FENKKKEAIEAETENTLNTLILKH----- 1087
Db 1744 LDLSRSLLGIDTEDALQGRNESCDSK-EHTSETTEGTPRHNDVHOICDQAOQDLNDI 1802
QY 1088 ---YKGLVK---YNGESSP-----LKTLSLSE----- 1107
Db 1803 EKITETGAVKPTGCSGSPDVTNVEPPGEDKTOGSSSCEISLFSGPNALVPMDFLGN 1862

Qy	1108	--ESIQEDNVALEENFKVLSKIEG----	KLKONLEKKKSLYSSGJH----	HLIEL	1157
Db	1863	EDHNLQLRKETSNEMLRLIHVIEDRDRKVESLNL----	EKKELDSKLHLOEVOYLMTKI	1918	
Qy	1158	KEVIKNNKVTGNSPSENNTDVNNALSESJ-----		1185	
Db	1919	EACIELEKIVGELKE--NSDLSEKLEFYSCDHQELLQVETSEGLNSDLEMHAKSSRED	1977		
Qy	1186	-----KKFLPECTDAVATVSESGSPTLEQSPKKPASHVGAESNTITTSO-	1233		
Db	1978	IGDNVAKVNDMSKRFPLDVENELSRIRSEKASIEHE-----	ALYLEADELEVVOETKEL	2029	
Qy	1232	--ANVDEVDVPIVPIRGSEEDYDILGOVVT-----	GEAVTPS-----VIDNTLSIE	1278	
Db	2030	CLEKDNENKOKVIVCL-----EEL-----	SVYTSERNQLRGELDWTSSKTTALDQISEMK	2081	
Qy	1279	-----NEVEVLYLKLPLAGVYRSCLKOLENNVMFNVNVDILNSRENNRENNVLE	1330		
Db	2082	EKTOLESHQSECLHCIOVAEAVKREKTEL-----	LOTLSSDVBELLKDKTHLOEKQS--LE	2137	
Qy	1331	SDLIPYDILTSNATVAKDPKFLNKEKR-----	DKFLSSYNTIKDSIDTIDINFMD	1381	
Db	2138	KD-----SQALSTLTKCELENQJOLNKEKELLYKSESJQARLESSEYKELNWSKALEALV	2194		
Qy	1382	VLGYYKI--LS-----	EKKXSOLDS-----IKKYINDKQGENKYLPLFNL	1420	
Db	2195	EKGFEALRLSSJQOEYHOLRGRIETKRLVRLEADEKKOLHIAKLERERENDSLDKOVEN	2254		
Qy	1421	IETLYUKVUNDKIDLEFVNLHAKVLANITYEKSANE-----	VKRIKEL-----NYUKT	1465	
Db	2255	LERELQSSSEENQELVILDAINSKAEVETLKQTOEENARSRLKIFELDLVTLRSEKENLTKQ	2314		
Qy	1466	IQDKLADPKKNNNFVGLADLSTDYNNHNNLTKFLTSGVWFENLAKTAVLSNLLDGLQGM	1525		
Db	2315	IOEQOGOLSE-----	LDKLLSS-----FKSLLEEKEQOAEI	2344	
Qy	1526	NISQHCQVKKQCPQNSGCFRHLNDREREKCSKLNTYKQDGKQVFN--	NPCNNENN	1578	
Db	2345	QIKSEKSTAVEMJON-----	QLKELNDVAVALSGDQETIKKATQESLDRPTEEBH	2393	

RESULT 3
 PCT-US95-16216-1
 : Sequence 1, Application PC/TUS9516216
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Yen, Timothy J.
 : APPLICANT: Rattner, Jerome B.
 : TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
 : TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
 : NUMBER OF SEQUENCES: 4
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Dann, Dorfman, Herrell and Skillman
 : STREET: 1601 Market Street Suite 720
 : CITY: Philadelphia
 : STATE: PA
 :
 : COUNTRY: USA
 : ZIP: 19103-2307
 :
 : COMPUTER READABLE FORM:
 :
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 :
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: PCT/US95/16216
 :
 : FILING DATE:
 :
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/353,700
 : FILING DATE: 09-DEC-1995
 : ATTORNEY/AGENT INFORMATION:
 :
 : NAME: Reed, Janet E.
 :
 : REGISTRATION NUMBER: 36,252

```

? TELECOMMUNICATION INFORMATION
? TELEPHONE: (215) 563-4100
?
? TELEFAX: (215) 563-4044
?
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 3248 amino acids
? type: amino acid
? STRANDEDNESS: not relevant
? TOPOLOGY: not relevant
? MOLECULE TYPE: protein
? HYPOTHEetical: NO
? ANTI-SENSE: NO
?
PCT-US95-16216-1

```

Query Match	4.2%	Score 356.5	DB 5	Length 3248
Best Local Similarity	18.9%	Pred. No. 8.2e-11		
Matches 350	Conservative 304	Mismatches 631	Indels 569	Gaps 79

Qy	125	YADIKHR--VANYULTTKEKLYAPOLFOGHTNHMLTLODNIHGEVULDG--YEELINELY	179
Dd	709	FSDOKHOKIEHNMCKLTKSOLT--GOVEDL--BHKQOLLN-----EIMDKRCQODHAEYE	761
Qy	180	KLNFYFDLLRAKLMDYCAN-----YCOIPFNKLIRANELDVLKLVFGYKRPDLNI	231
Dd	762	SLR---DLTKSDASLVTNEDHORSLLAFDOQRAHHNSP-----IIEQOSMSE	810
Qy	232	KDNVGMEDYTKKKKKTENINELIEBSKTTIDKNNAKTEBEKKKLVQAQYDLSYKQ	291
Dd	811	RSEGRLEADOSPXSALIONRVDSLEFSLES--OKONMSDLOCEBELVQIKOEIEENIMLK	869
Qy	292	LEBAN--NLISLREIDRTLKKN-----ENIKELLKIMBKNNPPRANGPNPLL	341
Dd	870	AEOMHOSVATTSORISKLOEDTSIAHONVVAETLSALEKKEKLO-----LL	916
Qy	342	DKNKIEHEKEIEIKETKTFNIDS-----LTTDLELEYULREKNKNIDISAKVETKE	396
Dd	917	--NKVLEDEQAEIOELKSNHLLLEDLSIKELQILSEFSLE--KKEMSSILSN--KREIE	971
Qy	397	STEPNETRNGVATYLRINDINNALNELNSFGDILNPFDTKERSKNITYDNKRFKFINET	456
Dd	972	LTOENG-----LTKETINASLOEKM--NLIO-----KSESPANYID--EREKSISEL	1014
Qy	457	KEKIKIKIE---KKIEBKKSYEDRS---KSLNDITKEVYKELLNELYDSKFNNNIDLNF	509
Dd	1015	SDQYKOEKLILLOCETGNAUYEDLOKYYAAOEKSKLECLINETSJLCEKRNKMLLOL	1074
Qy	510	EKKMGKRSYKVEKLYTGHNNFASYENSK--HNLEKLTALK--YMEDYSLRNTVYKKEKY	566
Dd	1075	KEAPAKENQOELTYL---AFABERNOLMLLETIYQOALYBSEMTDNQNNKSEMGGLKQ	1130
Qy	567	YKNLSIKIENIEETLVENIKDEEOLFEKTKTDE-----NKDPEKILYEVDI--	614
Dd	1131	EIMTLKEBQNNKMOKEVNDLLOENBOLMKWYKTHGEOQLNEBPRTNSYKERSERENQONF	1190
Qy	615	---VKVOYQAVLLMANKIDELKTLQILKINVELK-----HNTHV-----	649
Dd	1191	KPOMDLEVKETSIDSYNAOLVOLEAMRLMYELKLOESEKEXECLOHELOTJINGDETSLN	1250
Qy	650	-----FNSYKQEKQOPRYLLYLUKKEIDKLKFMKPYKESL	684
Dd	1251	QDMOSQETISGLKDCIEIDAERYTISGREHSTSONDANHLQCSLQTYMKNKLE--LEKICEI	1309
Qy	685	INEKKNKITKEGQSDNSEPTEGETIGQATTKRQOAG-----SALEGDSVQ	731
Dd	1310	LOAEKYEVLVELNDSRSECT-----ATRKAAEEVGLNBEVKLINDSDGLNGELIVE	1362
Qy	732	---AQAOBQKQAOAPVYUVPVPAKAQVPTPAPVYNKNTENYSKDYLEKLYEFLNTSYI	787
Dd	1363	DIPGEGEGEDORNEQHVSLL-----APRDESNS-----YEHUTLS--	1396
Qy	788	CHKIILVSHSTMMNEKL---KOYKTKREESKISSCDPLDLFLNTIONNIPWYMSFDSL-	843

```

Db 1397 -DKVQMHFAELQKPLSLQSEHKLHDQHCQMS-----KMSLEQTYVDSLK 1443
QY 844 -NNSLSOLFMEIYKEMWCNLYKLKDNKIKNLLEAKVSTSVKTLSSSSMQ----- 895
Db 1444 AENVLSTNLNLPQGDLYKEMQGLIEBLVPSLSSCVPDSSLSSLDSSFFYRALLEOT 1503
QY 896 -PLSLTPQDKPEVAN-----DQJSHSTNLNLSLFLFNISLSLG 933
Db 1504 GDMSLLSNLGAVSANQCSYDVEFCSSLOQTYVDSLAKENLVLSTLRFQGDLYKEMQGL 1563
QY 934 KNNKIYQEL-----IQKSSNFYKILK---DSDFYFN-ESTTNVVKKADI- 978
Db 1564 LEEGLVPSLSSCVPDSSSLSSLDSSFFYRALLEOTGMSLLSLNLEGVASANQCSYDVEF 1623
QY 979 -NSLNDSKRRK-----LEEDINKLTKTLOSPDLKYKYLKLERL 1018
Db 1624 CSSLOEENLTKETPPAPAKVEELESCEVYROSLEKLEKMSQGLMKKKEIQELPOL 1683
QY 1019 FD-----KKKTVCKYKQKIKLTLKLBQLESKLNS-----LNPRKHYLON 1058
Db 1684 LSSEROELDCRKQYLSNEQMOCKLTSVLEMSKLAEEKQTEQSLSELEVARLOLOG 1743
QY 1059 FSV-----FPNKKKKAETENTLENTKILKH----- 1087
Db 1744 LDLSRSSLLGIDEDAIQGRNESCDISK-EHTSETTEPTPHDHYQIDKDAQODLNDI 1802
QY 1088 ---YKGLVK---YNGGSSP-----LKTISE----- 1107
Db 1803 EKITETGAVKTCGSCGSGOSPDITVEPPGEKCTOGSSFCISELSFGNALVPRDFLGNO 1862
QY 1108 ---ESIQEDNVASLENFVKLSKLEG---KLKDNLEKKKLSYLSGLA---HLIAEL 1157
Db 1863 EDIHNLDLRVAKETSNENLRHLVIEDRDKVESLNL---EMKELSKLHLQEVOLMKI 1918
QY 1158 KEVAKNNNYGNSSENNTDVNNALSEY----- 1165
Db 1919 EACTELEKIVGELKKE-NSDLEKLEFYSCDHQELLORVETSEGLNSDLEMHAKSSRED 1977
QY 1186 -----KKFLPEGTDAVTVSESGSDPTEGSOQPKKPASTHYGAESNTITTSO- 1231
Db 1978 IGDVNAVAVNSMKREFLDVENELSRISSEKASIEHE-----ALYIEADLEVYQTEKL 2029
QY 1232 --NVDEVDVYIIVIFGESEEDYDGLQVVT-----GEAVTPS-----VIDNLSKIE 1278
Db 2030 CLKEDNEKQKIVICL---EEL---SVYTSERNOLRGELDTMSKKTALDQLSSEKM 2081
QY 1279 -----NEVEVYLKPLAGVYRSLSKKQLENNVTPVANNKDLINSFKRKREPKNVLE 1330
Db 2082 EKTQELSHQSECLHCIOVAEAEVKEKTEL---LQTLSSDVSELLKDKTHLOEKLOS-LE 2137
QY 1331 SDLIPYKDLTSSNVVADPYKFLNKKR-----DKFLSSVNYTKDSIDTDINFAND 1381
Db 2138 KD---SQALESITKCELENQALQNLKEKELVKESESLOARLSSEDEYKELNYSKALEALV 2194
QY 1382 VLGYYKI-LS-----EKYKSDLS-----IKKYINDKQGENEKLPLFELN 1420
Db 2195 EKGEFALRLSTOEBVHQLRGIEKLRVIREADEKQOLHIEKLERERENDSLKDVEN 2254
QY 1421 IETLYKVVNDKIDFVHLEKAVLNYTEKSNE-----VAKIKL-----NYLKT 1465
Db 2255 LERELQSEENQOELVILDAENSKAEVETLKTQIEEMARSLKIFELDLVTLRSEKENLTKO 2314
QY 1466 IODKLADPKKNNNEVGIADLSTVYNNHNLTKPLSTGTGVFENLAKTYLSNLDNLOGL 1525
Db 2315 IQEGOGOLSE-----LDKLLSS-----FKSLLEKEQDAEI 2344
QY 1526 NISOHQCVKQCPONSGCFRHLDEREBCKCLLNYKQSGDKCVENP-NPTCENN 1578
Db 2345 QIKESKTAEMELQN-----QKLELNEAVALACDQOEIMKATEQSUPPEIEEH 2393

```

```

: Sequence 8, Application US/08973462B
: Patent No. 6191270
: GENERAL INFORMATION:
: APPLICANT: DROITHE, PIERRE
: TITLE OF INVENTION: MALALIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
: FILE REFERENCE: 0660-0125-0 PCT
: CURRENT APPLICATION NUMBER: US/08/973,462B
: CURRENT FILING DATE: 1998-02-06
: EARLIER APPLICATION NUMBER: PCT/FR96/00894
: EARLIER FILING DATE: 1996-06-12
: EARLIER APPLICATION NUMBER: FR 95/07007
: NUMBER OF SEQ ID NOS: 29
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 8
: LENGTH: 1786
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

```

```

Query Match          4.2%; Score 351; DB 4; Length 1786;
Best Local Similarity 20.0%; Pred. No. 7,7e-11;
Matches 269; Conservative 243; Mismatches 437; Indels 398; Gaps 68;

```

```

QY 229 DNTDNGKAMEDYIKKKNKTENINELIES-KKTIKKNKATVEEKKKLYQAOYLDI 287
Db 757 ESVEENV---ESVAEENV---ESVAEENVEESVAENVEESVAPVEIEIAPSVESVAPSV 811
QY 288 YNKQLEE-----AHNLISYLEKRIQTLKKNNIKE-LLDKINETKNPPRANSQTPTML 341
Db 812 EESVAENVAATLSNLSNLLGLET-----EIKDSILNIEEYKE-----NVVTIL 860
QY 342 DKNKRIEHEKE---IKEIKTIKFNISLFTDPLELEYLRKKNRNIIDISAKVETRES 397
Db 861 E---NVEETTESVTTTNSLIEIQEN---TITNTIEEK-LEELHEV-LSAALENMQS 912
QY 398 TEPNEYNGVTPPLSYNDINNALNELNS--FGDLINPDYTKERSKINUYDNERRKLINE 455
Db 913 EEEKK-----EVIDVIEEVEEVAATLIEVEQAEEKSANT-----ITE 951
QY 456 IKEIKIEKKIESDKRSYEDRSK---SLNDITKEVEKILNELIYDSKPPNNIDLTNEK 511
Db 952 IFE---NLEENAVESNENVAEMLKELNETVNTVLDKVEIYEISGESLNNEMDKAFPE 1009
QY 512 MMGRYSYKVEKLTNHTNFASYENSKHNLKLTAKLYMEDYSLRNIVEREKELYRNLI 571
Db 1010 I-----FDNVKIQENL---LTGMPRSIE---TSIVQSEKVDLN-- 1044
QY 572 SKINELIETLVENIKKQEOULEK--KITDKENKPKDEKILEVSDIVAYQV-----QKV 622
Db 1045 ---ENVVSSILDNIEKKKEGLNKLLENISSTEGVOETVEHEQONVVDVAVPAMKQDPL 1101
QY 623 LNMKIDELKKTQOLILKNVELKNIHVNSYKQEN-----KQEPYILIVLKKETDKL 674
Db 1102 GILNEAGLKEMFNLBDV-----FKSBDVITYEIKIDEVVOKEVEKETYSII 1150
QY 675 KVPMPKVESLINEKKNIKTEGOSDNSEPTGEITQOATTKPQOAGSALLEGDSVOQA 734
Db 1151 EEMENETVLDVEEKKDL-TDKMIDAVEESIE-----ISSDS-KEET 1190
QY 735 QEQKQAOPPVPPVPEKKAQVPTPAPVNNKTEVNSLDLYLEKLYFLNYSYCHKTIYV 794
Db 1191 ESTDKKEDVSLVEEVOAD-----NDMDESV-----EKVLELKN----- 1224
QY 795 SHSTNMKILIKQYKITEEESKSLSSCDPLDLFLFIQNNIPVMSMPSLNSLOLFMEI 854
Db 1225 ---MEDELKMK-----DAVE-----INDITSKLEET 1247
QY 855 YE-KEMVCNLYKLKDNKIKNLLEAKKVSTSVKTLSSSSMQPLSLTPQDKPEVSANDPT 913

```

Db 1248 QELNEVADL--IKDMETKEL-----FKALSDE-----KEIIDAKDT 1285
 Qy 914 -----SH--STLNNSLKLFENILSGKKNITYOELIGKSS--ENFEYKILDSOTFY 963
 Db 1286 LEKVEEHDITTLTLDVEVL-----KDYEEKIEKVSOLKLEEDILE----- 1330
 Qy 964 NESFTNVKSAADINSUNDESKRKLLEEDINKLKTLOLSFDLYNFKYKLERFDKK 1023
 Db 1331 -----VKRIKLESEI--LEDYKEKLTETDLEKK--EIKDKPEKFEAEELKDLA 1302
 Qy 1024 TVGKYKMOIKRLTLKLEESKLSLNPKHVLNFSVFENKKKAEALTAETENTLENTKI 1083
 Db 1383 DI-----LKEVSSLVEEKKLEEVHELKE-----EVEHIIISGD-- 1416
 Qy 1084 LKRYKGLVITYNNESSPLKTLSEESIOTEDNVALENFK--VLSKLEGKIKDNLNLEK 1141
 Db 1417 --AHIKGL-----EEDDLREVVDLKGSIIDMLKGDMLG--DMKE 1453
 Qy 1142 KLSYLSGLHLILAEKVIKKNYTGNSPENNTDVNNALESYKFLPEGTDAVATVSE 1201
 Db 1454 SLEEVTKLGERVSLKDVLSA--LGMDEQMTKRKAQPKLEEVL-----LKEEYKE 1506
 Qy 1202 SGSDTLQS-----QPKPASTHVGAESENTITTSQNVDEVDVITVPIGSESEEDYD- 1254
 Db 1507 EPKKIKTKKRVFEDIKDEKPELVEYEMKDEDIEEDVEEDIEEDKVEDIDEDIDE 1566
 Qy 1255 DLQGVYGEAVTSPVINDILSKINEVEVLKPLAGYRSKQOLENNVTFNANVKDI 1314
 Db 1567 DIGE-----DKDEVIDILYOK--EKRIEKYAKK-----KLEKLEVEGVSGLKHVDEV 1614
 Qy 1315 LN--SRFNKR--ENFKNVLESDLPYKDLTSSNVVVDPYKFLNKKRDKRLSSYN--YIK 1369
 Db 1615 MKYQKIDKEVDKVSALSK-----NDVTN--VLKQNDPFSKVK--NFKYKIKYFAA 1665
 Qy 1370 DSIDTDFINPANDVIGY-----KILSEKYSDDLSDIKYINDROGENEKLPLN 1420
 Db 1666 PLSAVAFASVYVGFPTFLSFSSCVTIASSTYL--LSKVDTIN--KNKERPFYSF 1718
 Qy 1421 IETLYKTVNKKIDLFVYHLEKVLNITYEKSNSVEKIKELNLYKTIDOKLADFKNNNFV 1480
 Db 1719 VFDFLKNLK-----HYLOOMKEKFSK--EKNNNVI 1746
 Qy 1481 GIADLSTDYNNHNLTKFLSTGVFEN 1507
 Db 1747 EYTKAKKKGWQVYTKTEKTKYDKN 1773
 RESULT 5
 US-08-328-254-6
 ; Sequence 6, Application US/08328254
 ; Patent No. 5710022
 ; GENERAL INFORMATION:
 ; APPLICANT: Zhu, Xuellang
 ; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Campbell and Flores
 ; STREET: 4370 La Jolla Village Drive, Suite 700
 ; CITY: San Diego
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 92122
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/328,254
 ; FILING DATE: 24-OCT-1994
 ; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/141,239
 ; FILING DATE: 22-OCT-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Campbell, Cathryn A.
 ; REGISTRATION NUMBER: 31,815
 ; REFERENCE/DOCKET NUMBER: P-CJ 1191
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 535-9901
 ; TELEFAX: (619) 535-8949
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2482 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-328-254-6

Query Match 4.28; Score 350.5; DB 1; Length 2482;
 Best Local Similarity 19.84; Pred. No. 1.2e-10;
 Matches 353; Conservative 281; Mismatches 617; Indels 537; Gaps 79;

Qy 125 YADLKHR--VANYLLTIKELKYPOLFDTLNMHTLCDNIGHFKYLLIDG---YEEINELLY 179
 Db 77 FSDQKHQKEINMGLKTSQLT--GQVEDL--EHKLQLSN-----EIMDKRCYODLHAEYE 129
 Qy 180 KLNFFDILRAKLVDCAND-----YCOIPFNLIKIRANELDVLKLYFGYKRPIDNI 231
 Db 130 SLR--DLIKSKDASLYVNEHDSRLAFDQOPAMHHSFAN-----IIGEGSMPS 178
 Qy 232 KDNNKMDYIKKKKKTENENNELIEESKTDKNKNAKTEEEKKLYOAYDLSTNNQ 291
 Db 179 RSECRLEADQSPKSAIILQNRVDSLEFSLES--OKOMSDLOKQCEBLOIKGEIEBNLMK 237
 Qy 292 LEBAH--NLISVLEKRIIDLKKN-----ENIKELDKINEKNPPRANGTPTNLT 341
 Db 238 AEOHNSFVAETSORISKLOEDTSAHQNVVAETLSALENKEKELO-----LL 284
 Qy 342 DKNKIEHEKEIEIKTIFNIDS-----LFTDPLELEYLLEKKNKIDISAKVEIKE 396
 Db 285 --NDKVFTEQAEIOELKSNHLSLDSLSKEIQLTSETLSL--KKEMSIIISLN--KREIEE 339
 Qy 397 STEPNEYGVVTYPLSYVDINNALNELNSFGDLINPFDYTRPKSNITYDNERRKKTINEI 456
 Db 340 LTOENG-----TLKEINASLQEKM--NLIO-----KSESFANTID--EREKSISEL 382
 Qy 457 KEKIRIEK---KKIESDKSYEDRS--KSLNDITREYKLELNEIYDSKFNNDIDLTNF 509
 Db 383 SDQYKQEKLIILQCEETGNAVEDLSQYKKAQKNSKLECLLNECSTLCENRNELEQL 442
 Qy 510 EKMMGKRISYVEKLTTHNTFASTENSK--HNLEKLTKALK--YVEDYSLRNIVYEKELY 566
 Db 443 KEAFKHEHOEFLTKL---AFAEERNQNLMLLETVOALRSEMTDNOONSKSSEAGLKQ 498
 Qy 567 YKNLSIKENIEFTLVENIKKDEOLEFEKKTITKDE-----NKPDEKILEVSDI-- 614
 Db 499 EIMTLKEQNMQKEVNDLQENQOLMKVMKTHHCNLESEPIRNSYKERESEERNOCNF 558
 Qy 615 ---KVQYQKVLNNKIDELKKTOLILKNVEL-----HNHIV----- 649
 Db 559 KQPMDLVEKELISDSYNAQIVQLEAMLRNKELKLOESSEKECELOHELOTTROGLSTNLT 618
 Qy 650 -----PNSYKQENKOEPPYILVILKKEIDKLKVPMPKVESTL 684
 Db 619 QDMQSOEISGLKDEIDAEEKYISGPHELSTQNDNHLQCSLOTATNKNLINE--LEKICEI 677
 Qy 685 INEKNKNTKEGQSDNSPSEGETITGOATTKPGQOAG-----SALEGSQVQ 731
 Db 678 LOAEKVELVETLNDRSRCIT-----ATRKMAEYGVKLINEVKLINDSGLLHGLVELVE 730
 Qy 732 ---AQAOEQKQAPVAVPVPEAKAOPVTPPAPVNNKTEVNSKLDYLEKLYEFLNTSYI 787

```

Db 731 DDPGEGEOPNEQHPVSL-----ADLDESNS-----YEHLLTS-- 764
Qy 788 CHKYLIVSHSTMNKIL---KQYKTKREESKLS-----CDPLDLFNION 831
Db 765 -DKVEQHFALQEKFLSLQSEKILHDQHCQSSKMSLOTTYDSLKAELNLTSTLRN 823
Qy 832 -----NIPVWYMSFDSLNSLSQLFMEIYKEMVNCILYKDKDKIKNL 875
Db 824 FQGLVKEQMLGLEGVLPSLSSSCVDPDSSSLSGDSSFRLAL---LEQGDMSLSLNL 880
Qy 876 LEEKKVYSTVKTLSSSSMOPLSLTPDCKPEVSA---NDPTSHSTNLNSL-KLFENILS 931
Db 881 EGAVSANQCSVDEYFCSLQSEENLRKRETPAPAKVEELSLCEVYQSLSEKLEEKMS 940
Qy 932 LG--KKNKIYOELLGOKSSENFYEKILKDSPTFYNESFTNVPKADINDNESKRRK 989
Db 941 QGIMKNEI--QELQOLLSSEQOEIDCKOYLSENEOW-----QOKLJSTVLEMEKSL 992
Qy 990 LEEDINKLKTLOLSPDLNKKLLEKLPDKKRTVGKMKQIKKLTLLKQLESKLSL 1049
Db 993 AAEK-----KQEQSLSEL-EVARLOQGL-----DLSSRLIGIDTEDIAIGR 1035
Qy 1050 NNPKHVILONSVPFNKKKEAETENTLENTKILKH-----1087
Db 1036 N-----ESCDSK-EHTSETTEKPKHDVHQICDKDAQODLNDIEKIT 1078
Qy 1088 YKGLVK---YNGGESSP-----LKTLSR-----E 1108
Db 1079 ETGALKPGESESGQSDPTNTEPPGEDKTOGSSSECSISLSTSGRNALVPMDFLGNOEDIH 1138
Qy 1109 SIQTEDNVASLENKVLKSLGEG---KLKDLNLEKKLSTYLSGLH---HLIAELKEVI 1161
Db 1139 NLQLRVETSMENLRLLHVIEDRORKEVSSLN---EKKELDSKLHQEVLMLKRIEACI 1194
Qy 1162 KKNKVTGNSPENNTPVNNALEST-----1185
Db 1195 ELEKIVELKKE-NSDISEKLEFVSCDQOELLQVETSEGLNSDLEMHADKSSREDIGDN 1233
Qy 1186 -----KKFLPEGDVATVSESGSDTLEQSQPKKPASTHVGAESENTTTSO---NV 1233
Db 1254 VAKVNDSMKEFLVLENELKIRSEKASIEHE-----ALYELADLEVYQTEKLEK 1305
Qy 1234 DDEVDVYIIVPFIQSESEDDYDLOQVYT-----GEAVPSS---VIDNLSKIE--- 1278
Db 1306 DNEKKQKIYVCL---EEEL---SVVTSERNQRLGELDTMSKKTALDQLEKMKERTQ 1357
Qy 1279 ---NEVEVLYKPLAGVYRSLKQOLENNWTFVNVNKKDILNSFPNKRKNVLESOLI 1334
Db 1358 ELESQSECLHCIOVAEAVEKTEKEL---LOTLSSDVSELLKDKTHQELQOS--LEKD-- 1411
Qy 1335 PYKDLTSSNVVADPKYKFLNKEK---DKFLSSVYIKDSIDPTDINFANDVLYG 1385
Db 1412 -SQALSITKCELENOIOLNKEKELVKESLSQARLSESEYEKLNKALEALVGE 1470
Qy 1386 YKI--LS-----EKYKSDLS-----IKKYINDQGENEKYLPINNIETL 1424
Db 1471 FALSLSTOEVEHOLRGRGIEKLRVIRADEKQOLHIAEKLEKREKRENDSLDKVENELE 1530
Qy 1425 YKTVNDITDIFVHLEKAVLNTYEKSNE-----VKIKEL-----NYLKTQDK 1469
Db 1531 LQMSSENOELVILDAESKAEEVETLKTOIEEMARSLKAVFELDVLTLSEKENLKOI 1590
Qy 1470 LADPKRNNNEVGIALDSTDYNNHNLITKFLSTGWFENLAKTVLNSLLDGLQGLMINSQ 1529
Db 1591 QGQSLF-----LOKLSS-----FKSLLEKEQAEIQIKE 1620
Qy 1530 HQCVKQOCPONSGCFRHLDEREECKCLLNKQEGDKVEEND-NPTCNENN 1578
Db 1621 ESKTAVEMLQN-----OLKELNEAVNALCGDOEIMKATQESLOSDPTEBEH 1665

```

RESULT 6
US-09-104-324B-4

```

: Sequence 4, Application US/09104324B
: Patent No. 6232460
: GENERAL INFORMATION:
: APPLICANT: T recel, Ozlem; Sahin, Ugur; Pfreundschuh, Michael
: TITLE OF INVENTION: Methods For Diagnosis And Treating Cancers,
: AND METHODS FOR IDENTIFYING PATHOGENIC MARKERS IN A SAMPLE
: TITLE OF INVENTION: And Methods For Identifying Pathogenic Markers In A Sample
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fulbright & Jaworski LLP
: STREET: 666 Fifth Avenue
: CITY: New York City
: STATE: New York
: ZIP: 10103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
: COMPUTER: IBM
: OPERATING SYSTEM: PC-DOS
: SOFTWARE: Wordperfect
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/104, 324B
: FILING DATE: 25-June-1998
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/892, 702
: FILING DATE: 15-July-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Hanson, No. 6232460man D.
: REGISTRATION NUMBER: 30,946
: REFERENCE/DOCKET NUMBER: LUD 5491
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 318-3000
: TELEFAX: (212) 752-5958
: INFORMATION FOR SEQ. ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 976 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: US-09-104-324B-4

```

```

Query Match 3.98; Score 332; DB 4; Length 976;
Best Local Similarity 20.7%; Pred. No. 3.8e-10;
Matches 236; Conservative 204; Mismatches 354; Indels 344; Gaps 59;

Qy 196 CANYCQIPF---NLKIRANELD---VLKRVFGYKRPDLNKKNGKMEYIKKNNK-- 247
Db 41 CTEDLEPPFAKTNLSKNGENIDSDPALQKYNF---LP---VLEQVNSDCHVOEGLKDS 94

Qy 248 TIEN-----INLIESKKTIDKNNKATKEE-----EKKKLYQAO---YDLSI 287
Db 95 DLENSEGLSRVFSKLYKEAEK-IKKMWSTPAELRQKESKQENKRIIEORAKIOELOF 153

Qy 288 YNK---QLEFAHNLISVLEKRIIDTLKKNENIKELDKINEIKNPSPANSQNPNTLLDK 343
Db 154 GNEVYSKLEEG-----IQENKDLIKENNATRHLCMLKE-----TCARSAEK 196

Qy 344 NKRTIEHKEKEIKAKTIKFNIDSLF-----DPLELEYILREKKNKNT-----D 387
Db 197 TKKYEERETROYVMDLNNTIEKMITAFQELRYOAEENSRLEMEHFKKEDYEKIOHLEOE 256

Qy 388 ISAVETKES-----TEPNEYPNGVYTPLSYNDINNALNELNSFGDLINFDYTKE 438
Db 257 YKKEINDKEQVSLILOITFEKENKMDIFLL--ESRCKVNDLE-----EKTKL 305

Qy 439 PSKNITVDNERKKEFINIEKIEKIKIEKKKIESDKKSYED---RSKSLNDITKEYEKLLNE 494
Db 306 QSENLKOSIEKQHLTKLELEDIKVSLQSVSTOKALEEDLOIATKTIQQLTKEKETOEE 365

Qy 495 IYDSKFNNNIDLTNFE-----KMGKRIYSKYVEKULHNHTFASYENSKRN--LEKTLKA 546
Db 366 SNKRAAHSFVYTEFTTVCSLEELLRTEDQRLKNEKDQKILTMELQKSSSELEEMTKL 425

```

```

QY 547 L-KYMEDYSLRNIVKEKELKYKLNLSKIEIETLIVENIKDEQLEFKKIKTENKP 604
D 426 TNKREVELEELKVLGEKTELLEN-----KQEKIAEELKGTGEQELIGLQAREKVVH 479
QY 605 DEKLEVEDYKVOVAVLNNKIDELKKTOL-----LKNVLLKNNIHVPNSYKQENKQEP 661
D 480 D--LEI-OLVAITTSQOYSEKVKDL-KTELENEKLTNLTSTSC--NKLSTLENKE-- 529
QY 662 YLLIVLKEKIDKLVPFMPKVESLINEEKNKIKTGGOSDNSEPTGEGITGQATTKPGQQA 721
D 530 -----LQOETSDMTLELKNQOEDINNKKOEERMLKQIENLOETETOLRNE----- 575
QY 722 GSALEGSVOAOEOQOAPVPEVPEAKAQVTPPAPVNNKTEV--SKLDYLEKLYE 780
D 576 -----LEVREBELKQKR-----DEVCKIKDKSEENNNLRKQVENNKYITEELQO 620
QY 781 FLNITSYCHKIYLVSHSTNNEKILKQYIKTKEESKUSSCDPLDLENNIPVMSMF 840
D 621 -----ENKALKKKGTAEKQOLNVEI--KYNKLE--LEI----- 650
QY 841 DSLNNSLSOLEMET--YKEMVGNLVKKNKNDKI--KNLEPAKYSTSVKTLSSSSMQ 895
D 651 -----ESAKQKFGELTDYQKEL-----EDKKISENNLEVEYK--AKVIADKAV- 693
QY 896 PLSTLPQDKPEVSANDTJSHSTNLSNLSKLFENILSLGNKNKIYQELIGKQSENFYEKI 955
D 694 -----KLQKELDKRCQH-----KIAEWALMEKHKHQVDKILEEDSE----- 731
QY 956 LKSDTYNNESTFNFSKADINSLNDESKRRKLEEDINKLKTLOLSPDLNKKYKIKL 1015
D 732 -----LGLYKSKEQEOSSL-----RASLEIELSMLKAEL-----LSVKKOLEI 769
QY 1016 ERLPDKKRTVGYKQMOJKLKL-LKEQOLESKLNLSNKNHVLQNPVSFNNKKKELIET 1074
D 770 ER--EKE--KLKREKENTATLKEKKDKK-----TQTF----- 799
QY 1075 ENTLNTRKILKHYKGVKYNGESSPLKTLSEESIOTEDNYASLENKVKLSLEGKLD 1134
D 800 --LETFEITWK-----LDSKAVPSQIYSR-----NFTSVDH----- 829
QY 1135 NLNLEKKKLTSLSSGHLHLIAELKEVIAKKNVNGNSPS-----ENNTDVNNALESYK-- 1187
D 830 --GISKDKRYLMTSAKNTLSTPLP--KATYVKTPTKPLQORELNIPIESSKKR 882
QY 1188 -----FLPECTQVATVSSSGSDTLEQ--SQPKKPATVHGAES-----NTITT 1229
D 883 KMAFEDINDSDSETDLSWVSE--BETLKTLYRNNNPASHLCVTKPKKAPSSLTP 938

RESULT 7
US-08-687-080-51
Sequence 51, Application US/08687080
Patent No. 5965427
GENERAL INFORMATION:
APPLICANT: Gregory Dolganov
TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
NUMBER OF SEQUENCES: 175
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,080
FILING DATE: 17-JUL-1996
CLASSIFICATION: 435

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/592,126
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Shultz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 1312 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: TRANS. OF RAD50 CDNA (SEQ. 54), NT.
US-08-687-080-51

```

```

Query Match 3.7%; Score 308; DB 2; Length 1312;
Best Local Similarity 19.4%; Pred. No. 1e-08;
Matches 248; Conservative 214; Mismatches 453; Indels 364; Gaps 56;

```

```

QY 411 LSYNDNNALNELSGDGL--NPFDTKPEPSKNIYDNERKKFNEIKERIKIEKKI 467
D 84 LQFRDYN-----GELIIVQSMVCTQSKTEFTLEGVITRTFHGKXVLSSSCA 134
QY 468 ESKDKSYDRSKS--LNDITKETE-----KLNLIEYDSKFNNDILTNPEKMG 514
D 135 EIDREMISSLSGSAVLNNVIFCHQEODSNMPLESGKALKQKQFDEIFS-----A 182
QY 515 KRYSKYVEKLTNHNTPASYSKSNHLEKLTAKLYMEDY-----SLRIVYEKE--LKY 567
D 183 TRYIKALETLQVROTOGOKVETOME-----LKYKQYKKEACEIRQOITSKEQOLTS 237
QY 568 KNLISKIENEIETLIVENIKKDEQLEFKKIKTENKPEKILEVSDYKVOVK-----V 622
D 238 KEIYKSYENELDPLKNRKEIEHN--SKIMKLDNE-----IKALDSKKQMEKDNSELE 290
QY 623 LLMNKIDELKKTQLLKNNVELKNNIHVPNSYKQENKQEPYLLVKKEDIKLVMPKVE 682
D 291 EKMEKVFQGTDEOL-----NDLYHN-HQRTVREKERK-----LVQCHRELKLN--KES 336
QY 683 SLINDEKKNITEG-----QSDNSE-----PSTGEITGQ 712
D 337 RLNDKESKELLVEQRLQLOADRHOEHIRARDSLIOSLATQLELDGFERGPFSEQIKNF 396
QY 713 APTKPGQAGSA-----LEGDSVOAOEOQOAPVPEVPEAKAQVTPPAPVNNKTE 767
D 397 HKLVREOEGAKTANQLMNPAKETIKQO-----IDETRDKTGTGRIETLKSE 448
QY 768 NVSKLDYLEKLYEFLNITSYCHKIYLVSHSTNNEKILKQYIKTKEESKUSSCDPLDLF 827
D 449 ILSK-----KQNELKNV-----KYELQOLEGSSDRILELDELKAEKRELSKAEK----- 493
QY 828 NIQNNIPVMSFMSLNSLSQLEMEIYKEMVGNLVKKNNDK-----IKNLLEPAKK 881
D 494 -----NSNVETLKMEVTSIO-----NEKADLDRTLRKLDQEOE 527
QY 882 VSTSVKTLSSSMQPLSTLPQDKPEVSANDTJSHSTNLSNLSKLFENILSL-----GNK 936
D 528 LNH--HTTTRQMEMLTQDKADKQDOIRKIKSRHSDLTSLGFGFPNKKQLEMDLHSSK 585
QY 937 NIYQ-----ELIGQKSENFYEKILKSDTFYNESFTNFSKADINSLNDESK 986
D 586 EINOTRDLAKLNKELASSBQKNHNNELKRKE-----EQLSYEDKLDVCGSQDFESD 641

```

```

QY 987 RKLLEEDINKLK---TLQSFDLNKKYKRL-----ERLFDKRTYGVKYMQI 1032
D 642 LDRLEELIEKSSKORAMLAGATAVYSOFITQITDENOSCCVQCRVOTFELEQEVISDL 701
QY 1033 K-KITLKEQLESKLNINPKH-----VLQNSVFENKKKEAETAEENTLENTKI 1083
D 702 QSKRLAPADKLTSESELKKEKRDEMLGLVPMQSI--DLKEKEIPELRNKLQNVNR 759
QY 1084 LKHYKGLVKKYNGESSPLKLTSESIOTEDNVASLENFVYLSKLECKLKNLNEKKKL 1143
D 760 DIQRLKNDIE---BOETLLGTIMPEE--ESAKVCLDTVTIMEROMLKO---VERKI 809
QY 1144 SYLSSGHLHLTAELKEVETKNNKNTGNSPSENNTDNNALSESRYKFLPBGDTVAIVVSESG 1203
D 810 AQQAQAKLOGI--DLDRIVQYQVNOEKOEKQHKLDIVYSSKIELNRKLI----- 853
QY 1204 SDLEQSOFPKPASTHGAESNTTITS---QNVDEVDYIIVPIGESEEDYDLOGV 1259
D 854 QDOEOIOLHKTSTNELKSEKLIQISTNLORQOLEEOT---VELSTEVOISLYREIKD- 907
QY 1260 VTGEAVPSPVINDILSKTENEEVYLYKLPLAGVYRSUKKOLENNVMPFNVAVKDILNSRF 1319
D 908 -AKQVSP--LETLLEKFOE-----KEELINKKNTSNKTAODKLANDIK 948
QY 1320 NKRENFKNVLESDDLKYKDLTSSNVV--KDPYK-----FLNK-----EKRDKFLS 1363
D 949 EKVNINGYM-----KDI--ENVIOGKDYKKOKETELNKYIAQSECEKKEKINE 999
QY 1364 SYNTIKSDIPD-----INFANDVL-----GYTKLS-----EK 1392
D 1000 DMRLMRDIDIDYKIOERWLODNLTLRKRNELEKEVEERKOKHLEKMGOMQVLOKMSHOK 1059
QY 1393 YKSDLSIKKIKKOGENEKY-----LPLFNITETLYK-----TYND 1430
D 1060 LEEVINDIKRHNALGROKYESEIIFKKELEBPORDEEYREMAIYMTTELNVK 1119
QY 1431 KIDLEVLHLEKVLNVTYKSNVEKIKELNYLKTIOCKLADFPKNNNF----- 1479
D 1120 DLDIYVTLQDAIKMFH-----SMKMEIN--KIIRLWRSTYRGODIEYIEIRSDAE 1171
QY 1480 -VGADLSTDYNNHNLTK 1497
D 1172 NVSASDKRRNRYNVYMLK 1190

```

RESULT 8
 US-08-592-126-148
 ; Sequence 148, Application US/08592126
 ; Patent No. 5821091
 ; GENERAL INFORMATION:
 ; APPLICANT: Gregory Dolganov
 ; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
 ; TITLE OF INVENTION: Polypeptides
 ; NUMBER OF SEQUENCES: 151
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Dehlinger & Associates
 ; STREET: 350 Cambridge Avenue, Suite 250
 ; City: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94306
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/592,126
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Sholtz, Charles K.
 ; REGISTRATION NUMBER: 38,615

```

; REFERENCE/DOCKET NUMBER: 4600-0111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Rad50, pro-translation of SEQ ID NO:54
; US-08-592-126-148

Query Match 3.6%; Score 307; DB 2; Length 1312;
Best Local Similarity 19.4%; Pred. No. 1,2e-08;
Matches 248; Conservative 214; Mismatches 453; Indels 364; Gaps 56;

QY 411 LSYNDINNALNELSFGDLI--NPEDYKREPKNITYTDNRKKFINKIKIEKKKI 467
D 84 LQFRDVN-----GELLAVQSRNVCOTQSKTEKFTLEGVITRTKHGKVSLSKCA 134
QY 468 ESDKRSYEDRSKS--INDITKEYE-----KLNEIYDSKFPNNIDLTNFERKMG 514
D 135 EIDREMISSLGSAVLNNVIFCHQEDSNMPLSEKALKOKFDFLFS-----A 182
QY 515 KRYSKVEKLTHTNTPASYSNSKHNLEKTKALKYMEDY-----SLRNIVERE--LKY 567
D 183 TRYIKALETLQVROGOTOGOKVEKYOME---LKYLKQYKKEACEIRDQITSKEAQLTSS 237
QY 568 KNLISKIENELEIETLVENIKKDEBOLFEEKITKDENEKDEKLESDIYKVQYOK-----V 622
D 238 KEIYKSTENELDPLKRNKKEIEHNL--SKIMKLDNE---IKALDSRKKOMERKDNSELE 290
QY 623 LLMKRIKELKQTOLILKVELKHNHIVPNSYKQENKQEPYLIYKKEIDKLKFMPEVE 682
D 291 EKMEKVGQGTDEQL--NDLYHN--HQRTVREKERK-----LVQCHHELELN---KES 336
QY 683 SLINEKKNIKTEG-----QSDNSE-----PSTGEITGO 712
D 337 RLNQEKSELLVEQGRLOLQADRHQEHIRADSLIOSLATQLELDGERGFPFSEROIKNF 396
QY 713 ATRKPGQOAGSA-----LEGDSVOAQAOEQKQAPVPVPPVPEAKAVYPPPAVNNKTE 767
D 397 HKLYREROEGEAKTANQMLNDFAKETLTKQO-----IDETRODKTIGRIIEKSE 448
QY 768 NVSKLDYLEKLYEFLNNTSYICHKYLIVSHSTNNEKILKQYKITEEESKLSGCCDPLDLF 827
D 449 ILSK-----KQNELKNV-----KRELQQLBESSDRILELQDELKARELSKAK----- 493
QY 828 NIQNNIPVMSMFDLNSLSQLFMEIYEKEMVGNLYKLDKNDK-----IKNLEBAKK 881
D 494 -----NSNVETLMEVYSILO-----NEKADLDRLLRKLDQEMEQ 527
QY 882 VSTSVKTLTSSSMQPLSTTPQDKPEVGSANDFTSHNLNLSKLEFENILSL-----GKN 936
D 528 LNH--HTTTRTQEMELTKRADKDEQIRKIKSRHSDDELTSILGYPNKKOLEDMWLHRSK 585
QY 937 NIYO-----ELIGQKSENFEYKILKSDPTFYNESTFNPKSKADINDLNSDESK 986
D 586 EINTRORLAKLNKELASSDEQKHNHINELKRR-----EQLSSEEDLFLVCGSDPDESD 641
QY 987 RKLLEEDINKLK---TLQSFDLNKKYKRL-----ERLFDKRTYGVKYMQI 1032
D 642 LDRLEELIEKSSKORAMLAGATAVYSOFITQITDENOSCCVQCRVOTFELEQEVISDL 701
QY 1033 K-KITLKEQLESKLNINPKH-----VLQNSVFENKKKEAETAEENTLENTKI 1083
D 702 QSKRLAPADKLTSESELKKEKRDEMLGLVPMQSI--DLKEKEIPELRNKLQNVNR 759
QY 1084 LKHYKGLVKKYNGESSPLKLTSESIOTEDNVASLENFVYLSKLECKLKNLNEKKKL 1143

```



```

Db 760 DIOBKNDIE---EQETLLGTIMPEE---ESAKVCLDTVTIMERFQMLKD---VERKI 809
Oy 1144 SYLSSGLHLHLAELEKVNKNKYTGNSPSENNTDYNNALESKKRFLPECTDVAIVESG 1203
Db 810 AQAQAKLGI--DLDRIVQOVNOEKOEKQHKIDYVSKTELRKLI-----853
Oy 1204 SDTLEBOSPPKRPASTHVAESNTITTS---QNVDEVDVITVPFGESEEDYDLGOV 1259
Db 854 QDQOQIOHLKSTNELKSEKQIOISTNLORQOLEOT---VELSTEVOGLYEIND- 907
Oy 1260 VTGEAVTPSVIDNLSKIENEYVLYKPLAGVRSKQOLENNVMTFNVNKKDILNSRF 1319
Db 908 -AKEQVSP--LETTLEKFOE-----KEELLNKNTSNKIAQDKLNDIK 948
Oy 1320 NKREFRKVLSDLPYDRLSSNVV--KDPYK-----FLNK-----EKDKPLS 1363
Db 949 EKVKNHGYM-----KDI--ENYIQGKDYKQKETEELNKVIAQSECEKHEKINE 999
Oy 1364 SYNYSKDSIDTD-----INFANDYL-----GYKILS-----EK 1392
Db 1000 DMRIMRODIDQKIOERVLQDNLTKRKNELKEVEERKQHLKMGOMVLOMKSEHOK 1059
Oy 1393 YKSDLSIKKYINDKQGENEY-----LPLNNIETLYK-----TYND 1430
Db 1060 LEENIDNKRNNHLALGROKGYEEETIHKKELREPOFRDAEKYREMMIVARTELVKN 1119
Oy 1431 KIDLEVIHLEKVLNVTYEKSNVEYKIKELNLYKTIOKLADEKKNNF-----1479
Db 1120 DLDIYKTLDAIMKFH-----SKRMEIN--KIIRDMRSTYRGODIEYEIRSDADE 1171
Oy 1480 -VGJADLSTDYNNHNLTK 1497
Db 1172 NVSASDKRRNNYRVMLK 1190

```

```

RESULT 9
US-08-290-919-3
; Sequence 3, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DABRY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00367
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 21242/HCM/MJL/6BC8/

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /label= X
; OTHER INFORMATION: /note= "X = M and N, or N"
US-08-290-919-3

```

```

Query Match 3.6%; Score 301; DB 1; Length 53;
Best Local Similarity 100.0%; Pred. No. 4.7e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1570 NPTCTENNNGCGADAKCTEEDSGSNGKKTCECTKPDSPYLPFDGIFCSSN 1621
Db 2 NPTCTENNNGCGADAKCTEEDSGSNGKKTCECTKPDSPYLPFDGIFCSSN 53

```

```

RESULT 10
US-08-480-604A-6
; Sequence 6, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHAY, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:

```



```

STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLTA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-405-496A-6

```

Query Match 3.5%; Score 298.5; DB 2; Length 2710;
Best Local Similarity 20.4%; Pred. No. 8.1e-08;
Matches 339; Conservative 262; Mismatches 573; Indels 487; Gaps 90;

84 SGGVSAGSGVAGSGVAGSGSGNSRRTPSDSSDASADYADLKHRVRYLLTK--E 141
538 TGGSLSDNGVDENKNTALDKNYLLNKKIPSNVVEAGSKNYV-----HYIQLQDD 590
142 LKYPQLDL-----TNMLTLCDNHGF--YLIDGYEETNELLYKLFYDRLAKLND 194
591 ISYEATCNLSKPNKNSIITIQNNNESAKSYFLSDGESILE----- 633
195 VCAIDYCOIFNLKIRANEDLVKLKLVGY-----RKPLDNKIDNVGKMEYIK 243
634 ---NKY-RIPERLKNK---EKVAVTFIGHGKDFENFSEFRLKLSVDSLSNIISSFLDTIK 685
244 K--NKKITE-----NINE-----LIEESKKT-DKNNAATKEEEKKL 278
686 LDISPKNVVEVLLCGMNSYDFNVEETYPGKLILSIMDKITSTLPDVKNKSI-----TI 739
279 YQAOYDLSTYNKOLEE--AHN--LISVLEKRIIDLTKNENKELDKT--NKKNPAPNS 333
740 GANDYEVRINSEGRKELLAHSGKWINKEEALMSDLSKEYI--FFDSIDNKLK---AKS 793
334 GMP-----NTLDDKNNKIEEHEKEIKETAKTIKFNIDSLFTDPLELEYLREKN 383
794 KNIGGLASISEDIKTLDDASY-----SPOTKFTLNKIKLNIIESSIGDYI---YY----- 840
384 KNIDISAKVETKESTEPNEYPNGVYTPLSYNDINNALNELNSFGLDLPEDYTKEPSKNI 443
841 -----EKLEP-----VKNIIH--NSIDLDLIDEPMLNENVDEL 871

```

444 YTDNERKKFINEIKETIKIEKKIESDKSKYEDR-----SKSLNDI 484
872 Y---ELKK-LNNDEKYLISFEDISKNNSTIYRFINKNSGESYVETKEKIFSKYSEHI 927
485 TKREYKLEINEIYDSKFPNN---NIDLTFEKKMGGRYSKYEKKLTHNHTFASYSKHNLE 541
928 TKELSTIKNSIITDVGNGLDNIDLDHTSQVNTLMAFFIOSL-----IDYSSKNQVLN 981
542 KLTALK---YMEDYS--LRNIVEKELEKYKLNLSKIENE-----IETLV 582
982 DLSTSVKQVLAQLFSTGLNTIYDSIDL---VNLISNAVNTINVLPTEGIPVSTYL 1038
583 ENIK-----KDEQLPEKKITTKDENKPKDEKILEVDLVKQVOKVL----- 623
1039 DGINLGAIAIKELDEHDELKEL---EAKGVLAINNLSIAATVASINGAEVETIFL 1095
624 -----LNNKIDELKKTOLILKNVELKHNHIVPNSYKQENKQEPYLLVLKKKIDKLVM 678
1096 LPIGISAGIPSLVNNELLHD--KATSVNYFNHLSKSKYGP-----LKTEDD--KILV 1147
679 PKVESLINEEKKN-----IKTEGOSDN-----SEPTGEGITGOATTKP 717
1148 PIDDLVISEIDFNNNNSIKLGTGNILAMEGSGHVTGTGNIDHFFSSPSISHI-----P 1200
718 GQAGSALLEGSOQAQOQKQAPVPVPVPEAKAOY---PTPPAPVNNKTEN--VSKL 772
1201 SLSTYSALGITEMLDPSK-----IMLPNAPSRYFMETGAVPQLNSLENDGTRLL 1253
773 DYLEKLY--EFLNTSYICHKYLIVSHSTMEKILKOYKITEEESKLSGCDPLDLELNIQ 830
1254 DSIRDLYPGKRYMRYFAFDVAITTLKPYEDTNKIKLKDQFNFLM--PITTTMEIR 1310
831 NNIPVWYSMEDSLNNSLSQLFMEIYKEMVNCNLYKLKDNKIKMLLEBAKVSTSVKTL 890
1311 NKL--SYS-FDGAGGYVS-LLLSSY--PISTNIMLSKDDLMIFNIDVEPREISTENGIRK 1364
891 SSSMQPLSLTPQDKPEVSANDTSHSTNLNLSLTFENIIS--LGKNNKYQELIGOKS 947
1365 KG-----KLNDVLSKIDINKKLI-----IGNOT 1389
948 SENFEKILKSDPTFYNESEFTNFKSRADDINSINDESKRRKLEEDINKKTLQLSFDL 1007
1390 ID--FSGDIDKKQYI-----PLTCELDDKISLITE-----INLVAKS----- 1425
1008 YNKYKLELERLPDKKKYVYKAKMOIKLTLLEKDESKLNSLNPKNHYLQFVFFPKKK 1067
1426 ---YSLLSIG--DKNYLISNLSNPTIEKINTL--GLDS-----KNIAVNYTDESNNKY 1470
1068 EAELAEFTENTLENTKILKKH---KGLVKKYNGESSPLKLTSESIQTEEDNVASLENFK 1123
1471 FGAISKT-----SOKSITI-HKKKDSKNLTERYN--DSTLENSKDFIADIEDIVFMKDD-- 1520
1124 VLSKLEGLKLDNMLEKK---KLSYLS-----SGLH---HLIELKEVIRKNNKYTGNSP 1171
1521 -INTTGYKYVDNNTDKSIDRSISLVSKNQYKANGVILNSESYSYLDYFVKNSGHHNTS 1579
1172 SENNTDVNNALESTKFLPESTDVA---TVSESQSDTLE--OSQPKKPASTVGAESN 1225
1580 NFMNIFLDN--ISFMRKLGFEININVIDKYFTLVGKTYNGVFEFTCDNNKKNIDIFYGEWKT 1638
1226 TITTSQNVDEVDVIVPIPIGSESEEDYDLDGVVTVGAVPYSYIDNLSIKIENEVEYLY 1285
1639 SSKSTIPISGGRNVVEPIINPD-----TGEDISTS-----LDSYE--- 1676
1286 LKPLAGVYRSLLK-----OLENNVTFENVVKD-----TLN--SRKNRKNFNKYL 1331
1677 --PLYGIDRYINKVLAPDUTSLININTNYSNMEYPEIIVLAMPNPFKKVNLN--LDS 1732
1332 DLIPK-----DLTSSNYYVVKDPYKFLNKKRDKPFLS--SYNTYKOSID--TOINFANDVL 1383
1733 SSFEYKWSGTEGDFLVAVYEESSNKKILQKIRIGKILSNOSFMRKMSIDFKIKLS--L 1790
```



```

QY 1175 NTVDNNALES--YKFLPECTDVAIVVSESGDYLEOSQPKKPASTHVAGESNTITTSQN 1232
Db 2585 -VNLSALEGLFPAKSVTGDFAF-----KNTTLPT-- 2616
QY 1233 VDDEVDDVIVPIGESEEDYDGLGVVTVGEAVT--PSVIDNLSKLE-----NE 1280
Db 2617 --DGVSSILLIPPVYKE-----DIFHLFCGSKSTYKPKNKNTSIALIHIHISNRNIHG 2669
QY 1281 YEVYLLPLAGVYSLKQKLENNVNFENVNYKDL-----NSRPFKRE 1323
Db 2670 CDFLYLENQNDIAISNNNNNSYSIFTHNKNTENNLCIDISLIPVTGICPKPKKLNPT 2729
QY 1324 NFKNV--LESDDLIPYKDLTSSNY-----VVKDPYKLEKEDKFLSSNYI-- 1368
Db 2730 CFDEYVYVKEQEDVPSTITADKYNTFSKDKIGITLKNAISINNDEKD--NTYTYLILP 2786
QY 1369 ----KDSIDPIDINANDVLYGTYKILSEKYSDDLISIKKYINDKOGENE--KY-----LP 1416
Db 2787 EKFEEDLIDTRKVLACTGDNKYIIMKIEKSTMDKIK--IDEKTKICKICKYDVTTKVA 2844
QY 1417 FLNNIETLYKTV-----NDKI--DLFY--THLEKAVLNY 1446
Db 2845 TCEIIDIIDSSVLEKHEHTVYHYSITLSKMDKLIITYPNEKTHFENFVNPPLNDKVL-Y 2903
QY 1447 TYEKS--NVE-----VKIKELNYLKTIDKDLADFKKNNFVGIADLSTDYN 1490
Db 2904 NYNKPINIHLPGALITTDIYDTRKTIQ--YLIRIPPY--HKDILH-----SLEFN 2952
QY 1491 HNNLLTFSLTGAVFENLAKTVLSNLDGNIQGLNLSQ-----HOCYKQKOPNSGCFR 1545
Db 2953 NSLSLTK-QONONIIYGVAKI-----FIHINOYKEIHG--DFTGKYS 2993
QY 1546 HLDRECKCLNLYKQEGDKCVENPPTCNENNG 1579
Db 2994 H-----LFTYSK--KPLPNDDOICNVITG 3015

RESULT 13
PCT-US93-07261-11
; Sequence 11, Application PC/TUS9307261
; GENERAL INFORMATION: PFEIMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
; TITLE OF INVENTION:
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: John H. C. Blasdale
; STREET: One Giralda Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07940-1000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07261
; FILING DATE: 19930805
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,531
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blasdale, John H. C.
; REGISTRATION NUMBER: 31,895
; REFERENCE/DOCKET NUMBER: DX0288K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-822-7398
; TELEFAX: 201-822-7039
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1588 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: protein
PCT-US93-07261-11

Query Match 3.5%; Score 296; DB 5; Length 1588;
Best Local Similarity 20.3%; Pred. No. 5,7e-08;
Matches 308; Conservative 248; Mismatches 601; Indels 362; Gaps 69;

QY 243 KKKNTKIENTINE--LIESKTIIDKNNKATPEEKKKLYQAOYDLSYNNKLEFAHNLS 300
Db 12 EKNKARNALKEKKLKQKQKNDQAKADLTAKESQD-----SSSEKSLKEKVNGEA 62
QY 301 VLEK-RIDTEKK--NENIKELDKINEINPPANGTPTNLLDKNKITEHEKEIK 355
Db 63 LKEKENNETLKKKLELQNGKEKEKNKIKDNNDLAKKGNQKDKKIVPKKPESEVDLK 122
QY 356 EIAATIKFNISLFTDPLEEYLYREKKN--IDISAKVETESTEPNEDYNGVYTPLSYN 414
Db 123 EMELKEKEFIQHLKDYEE--RREKRRNWLRSLRDKLREIEOLEKLVNA----- 170
QY 415 DINNALNELNSFGDLINPFDTKPESKNIYTDNKRKKFINKIKIKIEKKIESDKSY 474
Db 171 QLESAINELKERRASRRPMYKQMGKDEYDEWIKYIDQAKKNGTKDEIKQGDGY 230
QY 475 ED-----RSKSLNDITKEYEKLEINEIYDSKFNNNIDLTNFEKMGKRYSYKEVLT 525
Db 231 EEIFETKFGYKRENALGEL-DEYEE-----REK--KRYLAK----- 264
QY 526 HNNTFASYENSKHNLKLTALK-----YMEDYSLRNIVEKELKYKNLISKIENEIET 580
Db 265 -----EDEGDLKDYEEKLEETGYGFRKFPPTRLIVARK--RNKQK----- 305
QY 581 LVENIKQDEDLFEKKKITDKENKPKDEKILEYSD-----IYVQYQYVLYNMKIDELKKTQ 635
Db 306 --KLKEDK--EKKLAAEPDEKKIKLKDSDDKVVAVN----- 341
QY 636 LILKNVELKHNHIVNPSYKQENKQEPYLL-----IVLKEIKDLKVFMPKVESLINEE 688
Db 342 -----KNKSPFDKFRAPDKRTMEFYRLSELFPVPRND-NELAVCGSGMSQKNGK 392
QY 689 KKNITKQSQDSNESPSTGEITGOATTKPGQOAGSALBGDSVQ-----AQA 734
Db 393 KL-----KSTFNPKRRRNKKLKERKMQELHKFKKRYKQKLE 431
QY 735 QEOKAQPPVPVPEAKQVTPRPAPVNNKTENV--KLDYLEKLYEFLNTSYCH 789
Db 432 REKRENPDGEPPLNTPETHV--IRPSDLMQGENKSGHPKPYPTGKLEY-ESHVSK 487
QY 790 KYILVSHSTNNE-----KILQYKITEEESKLSQCDPLDLFNIONNIPVMSFDS 842
Db 488 DYOL-EHEPPTKLPREYKGGHVSREYQDHPPTKLPRE-----KGHVSREYQDNE 538
QY 843 LNSLSQLEMEIYKEMVCNLYKLKDNKIKNLLLEAKKYSTVYKLSSSSQPLSLTPQ 902
Db 539 VRDELPE-----YKGGHVSREYQD-DNEGSPSTLKEYQDTELAKKIDITNKPHE--SVDEX 590
QY 903 DKPEVSANDOTSHSTNUNNSLKLFEENISLQKNKI-----QQL-----IGQ 945
Db 591 DQTELAKKIDITNKP--HESVDEYDQ--SELAKKIDITNKPHEVSDEYDQTELAKKREYTN 647
QY 946 KSENFYEKILKSDDTFYNESFTNFKSKADDINSLNDESKRK-----KLEIDINKLKT- 1000
Db 648 KPHENLEE--YNETDLAKGKEVTNKPHEVSDEYDQ--SELAKKIDITNKPHEVSDEYDQTE 704
QY 1001 LQSLFDLYNKYKTLERLFDKKTGVKQWIKKLTLLKQLESLK--NSLNNPKHYVLQ 1057
Db 705 LAKGEVJTNKARENILEYNETDLAKGK-EVTNKAARENILEYNETDLAKGKEVTNKAH--E 761
QY 1058 NFSVFPNKKKAELAEENTLENTKILKHKGLVKKYNGSSPLKLSSEISTEDNYA 1117
Db 762 NLEEY--NETDLA-----KGKEVTNKAARENILEYNETDLAKGKEVTNKA-----HE 805
QY 1118 SLENPKVLSKLEGLKLDNL--NLEKKKLSYSSGLHHLIAELKEVIKNNKNTGNSPEN 1174

```

```

Db 806 NLEEVNETDLAKGKEVTKARENLEEVNETDLAKG-----KEVT-----NKAREN 850
Qy 1175 NTDVNNALLESYKFF-LPEGTIVATVVSSESDTLEOSOPKRPASHVCAESNTITTSQNV 1233
Db 851 -----LEEYNETDLAKGKEVTKARENLEEVNETDLAKGKEVTKAREN--LEEYNE 900
Qy 1234 DDEVDVIVIFEESEEDYDGLQVVTGEAVTPSVIDNIILSKITENEX---EVLTKPLA 1290
Db 901 TDLAKGKEVTKARENLEEVNET--DLAKGKEVTKARENLEEVKEVTKARENLEONKQSD 959
Qy 1291 GYVSLK---KOLEN-----NVMTFNNAVKDILNSFRKRENFKN-----VL 1329
Db 960 GLKENAELKNEKLNKSGDLKENAELKNEKLNKSGDLKENAELKNEKLNKSGDLKE 1019
Qy 1330 ESDLIPIKDL-TSSNYVVKDYPKFLNKEKRFSS-----YNTKISIDIDI 1376
Db 1020 ENAELKNEKLNKSGDLKENAELKNEKLNKSGDLKENAELKNEKLNKSGDLKE 1079
Qy 1377 NFANDVL---GYKILSEKYSGLDSIKKYINDROGENEKLPLFNIETLYKTIVND-KI 1432
Db 1080 ELKKEKLNKSGDLKENAELKNEKLNKSGDLKENAELKNEKLNKSGDLKE 1134
Qy 1433 DLEVIHLEAKYLYTYEKSNEVRIKEL-----NYLKT--IQDKLADFK- 1474
Db 1135 NAEKLNKNEKLNKSGDLKENAELKNEKLNKSGDLKENYVYTNNDLKNNDIKNK 1194
Qy 1475 -KNNNFAGIADLSTDYNNHNNLLTFLSTGWFENLAKTVLSN--LLDGNLOGMLNISHO 1531
Db 1195 MNKNEKLNKDSNMKNEKLNKDLN---EDMKNEKLNKDIKNKDLSTIGMEQON 1250
Qy 1532 CVKQCP---QNSGCFRHLDRECEKCLLYKQEGDKCVENPN-PTCENNNKGCDDADAK 1586
Db 1251 TGLKNTPSKQGNQNTGLKNTPRQONTGLKNTPSGQONTGLKNTPSGQONTGLKNTPN 1310
Qy 1587 CTEEDSG-----SNGKIT 1600
Db 1311 EROQNTGLKNTPSGQONT 1329

RESULT 14
PCT-US93-07261-16
; Sequence 16, Application PC/TUS9307261
; GENERAL INFORMATION:
; TITLE OF INVENTION: PFEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John H. C. Blasdale
; STREET: One Giralda Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07940-1000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07261
; FILING DATE: 19930805
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,531
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blasdale, John H. C.
; REGISTRATION NUMBER: 31,895
; REFERENCE/DOCKET NUMBER: DX0288K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-822-7398
; TELEFAX: 201-822-7039
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 1663 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; STRAIN: Malayan Camp
; PCT-US93-07261-16

Query Match 3.5%; Score 296; DB 5; Length 1663;
Best Local Similarity 20.3%; Pred. No. 6e-08;
Matches 308; Conservative 248; Mismatches 601; Indels 362; Gaps 69;

Qy 243 KNNKTIENTINE--LIESKTTIDKNNATYEEKKLYQAYDLSYNKOLEAHNLIS 300
Db 12 EKNEKARNALKEKKIKDQKNDAQKADLTFRKESOD-----SSSEKLEKRVNGEA 62
Qy 301 VLEK-RIDTLK---NENIKELDKINEIKNPPANGNTPTNLLDKKKIEEHEKEIK 355
Db 63 LKEKENETLKKLELONKEKEENKIKDNNDALKNKGNKDKKIVPKKPESEVEDLK 122
Qy 356 EIAITIKFNIDSLFTDPLEEYVLRKNKN-IDISAVETKESTEPNEYPNGVYPLSYN 414
Db 123 EMELEKEFEIKQHLKDYE---RKEKRNWILRSLRDKLREIOLKLA----- 170
Qy 415 DINNALNLSFGDLINPDYTKESKNITDNERKKFINKIKIEKKIESDKSY 474
Db 171 QLESAINELKERRASRRMYKQGMKDEYDEWIKRYDQAKNGCTKDEIKDKGDY 230
Qy 475 ED-----RSKSLNDITKEYEKLNEIYDSKFNNNIDLTNEKMKGRYSYVEKLT 525
Db 231 EEIYETKFGYKRENAIGEL-DEYEE-----REK--KRYLYK----- 264
Qy 526 HNTFASTENSKNHLKLTALK-----YMDYSLRNIVKELKYKLNLSKITEIET 580
Db 265 -----EDGEGLKDYEEKLEETGYGFRKFPPTTLVARK-----RNKEOK----- 305
Qy 581 LVENIKKDEQLFEKKIKITDKENKPKDEKILEYSD-----IYVYQYQVLYNMKIDELKKTQ 635
Db 306 ---KLKEDK---EKKLIAEPPDEKKIKLKSDDKVYVFN----- 341
Qy 636 LILKNEKLNHIVPNSYKQENKOEPPYL-----YLVKEIDKLKVFMPKVESLINEE 688
Db 342 -----KNNSPDKFRAPDKKRTMFYRLSELPIYPRKD-NELAVGSGMSKYNKG 392
Qy 689 KKNITGSGDSEPSSTEGEITGOATTKPGQAGSALGDSVQ-----AQA 734
Db 393 KL-----KSTFNPFKRRNKKLKERKMOELHFKKYNKKYQKLE 431
Qy 735 QEOKQAPVPVPEAKAQPVRPAVNNKTEVNS-----KLDTLYKIEFLNTSYICH 789
Db 432 REKRENDGEPLNPELHV---IRPSDIMGKNGSKGHPKTYOPTGLKEY-BESHVSK 487
Qy 790 KYILVSHSTINE-----KILKQYKITEESKLSGCDPLDLFNIONNIPVMSFMS 842
Db 488 DYOL-EHEPPTKLPEYKRGHVSREYQLDHEPPTKLPEY-----KGHSREYQLDNE 538
Qy 843 LNNLSQLPMEIYKEMWNCNLKLDNDKIKNLLEBAKKYSTVKTSSSSMOPLSLTPQ 902
Db 539 VRDLPE---YEKGHVSREYQL-DNEGSPSTLKEYDQTELAKGDKDITNKPHE--SYDEY 590
Qy 903 DKPEVSANDPSSHSTNNLSKLLENILSLGKNKI-----YQEL-----IGQ 945
Db 591 DQTELAKGDKDITNKP--HESVDEYDQ--SELAGKADITNKPHEVSDEYDQTELAKGKEVTN 647
Qy 946 KSENEFEKILKSDTFYNESEFTNPFVSKADDINSKADSKK-----KLEIDINKLKT- 1000
Db 648 KPHENLE--YNETDLAKGKEVTKPHESVDEYDQ--SELAGKADITNKPHEVSDEYDQTE 704
Qy 1001 LQLSFDLYNKKYKLEKRLFKKKTGVKKYKQIKKLLTLKQLESKL---NSLNNPKHYIQ 1057
Db 705 LAKGKEVTKARENLEEVNETDLAKGK-EVTNKAARENLEEVNETDLAKGKEVTKAH--E 761

```

QY 1058 NFSVFNNKKKEAEIATENTLENTKILKHYKGLVKYYNGESSPLKLTSEESIOTEDNYA 1117
 Db 762 NLEEY-----NETDLA-----KGKEVINKAHENLEEVNETDLAKGEVTNKA-----HE 805
 QY 1118 SLENFKVLSKLEGLKLNL---NLEKKKLSYSSGLHLLAELKEVINKKMYGNSPSEN 1174
 Db 806 NLEEVNETDLAKGEVINKAHENLEEVNETDLAKG-----KEVT-----NKAREN 850
 QY 1175 NTDVNNMLESYKRF-LREGTDVATVSESSDITLQSQPKKPATPHCAESNTITTSQNV 1233
 Db 851 -----LEEYNETDLAKGEVINKAHENLEEVNETDLAKGEVTNKAAREN---LEEYNE 900
 QY 1234 DDEVDVYIPIFGESEEDYDDLQVVTGAVTPSVIDNITLSIENEX---EVLTKPLA 1290
 Db 901 TDLAKGEVINKAHENLEEVNET-DLAKGEVINKAHENLEEVNETDLAKGEVTNKAAREN---LEEYNE 959
 QY 1291 GYVRSK----KOLEN-----NVMTFNVVVKDILNSRFNKRENFKN-----VL 1329
 Db 960 GLKENAELKRELKNGSDGLKENAELKRELKNGSDGLKENAELKRELKNGSDGLK 1019
 QY 1330 ESDLIPIKRD-TSSNVYVKDPIKFLNKKRDKPLSS-----YNYIKSIDIDI 1376
 Db 1020 ENAELKRELKNGSDGLKENAELKRELKNGSDGLKENAELKRELKNGSDGLK 1079
 QY 1377 NFANDVL---GYRKILSEKYSDDLKIKYINDKGENEKYLPFLNIEETLYKTVD-KI 1432
 Db 1080 ELKKELQONG-----SEGKLENAEQNKRELQONGSEGLKENAELKRELKNGSDGLKE 1134
 QY 1433 DLEVILHLEAKVNTYKRSNVEKIKEL-----NYLKT--IODKLADK- 1474
 Db 1135 NAEIKRELKNGSDGLKENAELKRELKNGSDGLKENYVNTNDLKNQDIONKDSNKD 1194
 QY 1475 -KNNFPGIADLSDYVHNNILTFELSTGMVFEWLAKTVLSN--LLDGNLOGMLNISHO 1531
 Db 1195 MKNELKNDKDISNMDKMKELKNDLSN---EDMKKRELKNDKIRNKDLSISNMEQON 1250
 QY 1532 CVKROCP-----QNSGCFRHLDERECKCLILNYKQEDKCVENPN-PTCENENNGGCDADAK 1586
 Db 1251 TGLKNTSKQOQNTGLKNTPREQOQNTGLKNTPSGOQOQNTGLKNTPSGOQOQNTGLKNTPN 1310
 QY 1587 CTEEDSG-----SNGKKIT 1600
 Db 1311 EROQNTGLKNTPSGOQOQNT 1329

RESULT 15
 US-08-446-855A-2
 ; Sequence 2, Application US/08446855A
 ; Patent No. 5849573
 ; GENERAL INFORMATION:
 ; APPLICANT: Stewart, Thomas S
 ; APPLICANT: Flores, Maria V
 ; APPLICANT: O'Sullivan, William J
 ; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
 ; TITLE OF INVENTION: phosphate synthetase II
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Nixon & Vanderhye PC
 ; STREET: 1100 NO. 5849573th Giebe Road, 8th Floor
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: USA
 ; ZIP: 22201-4714
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.24
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/446,855A
 ; FILING DATE: 06-Jul-1995
 ; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
 NAME: Mitchard, Leonard C
 REGISTRATION NUMBER: 29,009
 REFERENCE/DOCKET NUMBER: 47-80
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-816-4000
 TELEFAX: 703-816-4100
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2391 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-446-855A-2

Query Match 3.5%; Score 293.5; DB 2; Length 2391;
 Best Local Similarity 19.1%; Pred. No. 1.3e-07;
 Matches 323; Conservative 252; Mismatches 543; Indels 577; Gaps 83;

QY 171 YEEIN-----ELVYKLNFPDLIRAKLNDVCANDYCOIPFNLIKTRANDVLKLYF 222
 Db 232 YKEINLFDPCNIDFLKYCNHFIRVI--KLNN-----ITYNK-----NKEEF 272
 QY 223 GYRRPLDNIKQNVGMEDYIKKNTTINELIEESKKTIDKKNATKEEKKKLYQAQ 282
 Db 273 NY-----TNEMITNDSSMEDHNEINGSISFNMC--PSISFDKS-----ESKNV----- 316
 QY 283 YDLSTYKOLEEAHNLISVLEKRIIDTLKKNENIKELDKINEIKNPANSGNTPNLLD 342
 Db 317 ---INHLLDKNMLTISSEYELKDL-HNCNFSNSDK-----NDSFF 355
 QY 343 KKKIEHEKEIKELAKTIK---NIDSLTPDLELEYLREKKNKIDISAKVETKESTE 399
 Db 356 KLYGCEYDKYILDEENASPHYNNVD-----EYGYDVNKKNTNLSNKKIQONNNE 408
 QY 400 PNEYVGVTVPLSYDINDNALLNEL-----NSFGDLINFDYTKPSKNIYTD-NERK 450
 Db 409 NKK-----NKKNNNNNEVDYIKKDEDNVNNSKVPYSQYNNNANQNNHEHFNILNN 457
 QY 451 KEINEIKERIK-----IEKKIESDKSYEDRSKSLNDITKEYEKLNEIYDSKFNNN 503
 Db 458 DYSTYIRKKMKNEEFLNVNKKRYDHRK-----IIVYVCGIKNS 498
 QY 504 IDLTFEKKMGKRYSYKVEKITLHNTFASYE-----NSKHLLEKTLKALTYMEDYSLNT 558
 Db 499 IIKMLIRHGMPLPLYIIVPYYNFNHIDYDAVLLSNQPGPKCDPLIKMLKDSLTENK 558
 QY 559 VV-----EKKLYYKMLISKIENETJEVNIKKDEO-----LEPKK 596
 Db 559 IIFGICLGNOLLGISLGDYKMKRYGNR--GVNOPYQLVDNICYITTSQHHGCLKRRS 615
 QY 597 ITKDE-----NRPDEKILEVSD---IVKQYQVYLLMMKIDELKTKQTLILNVELKH 645
 Db 616 ILKRELALSYINNDKSIDIGISHKNGRFYSVQFHP---EGNNGPETSLSLFFNELL- 669
 QY 646 NIHPNYSKQENKQEPYVLIYVKEI----- 671
 Db 670 --DIFNKKQYREYLVGNIYIIRKRVLLLSGGLCIGAGEFDYSQVAILSKLECGIYV 727
 QY 672 -----DKLVMPK---VESLINEKKN--IKTEQDSNSSEPTGE 708
 Db 728 ILVFNATVQTSKGLDKV-YPLPVNCFEVEKTIKKRPPFLICTG- 774
 QY 709 ITGQATTKPG-----QAGSALEGDSVQAQ---AOEQQAOPVPVYPVE 750
 Db 775 --GGTALNCALMDQKVKLKKNNQOCLGTSLESIRITENRFLFKELKEIMERI-APYGS 831
 QY 751 AKAOVPPPPAPVNNKTENVSKLDY--LEKLYER--LNTSYI-----CHKYILVSH 796
 Db 832 AK-----NVNOAIDIANKIGYPIIVRTTFSLGGLNSSFINNEEELIECKNRIFL-- 880

QY 797 STME-----KILK-----QKITEEES-----KLSCDPLDLFIQNNIPWYSMD 841
 Db 881 QTDNEIFIDKSLQCKEIEYELLRNNKNCIAICMENIDPL---GIHTGDSIVAPSQ 936
 QY 842 SLNNSLSOLEMEIYERENY-----CNL-YKLDNDKIKNLEBAKVSTSVKTLSSSS 893
 Db 937 TLSNVEYKFFREJIALKYTHLNIIGECNIOFGINPOTGEYCIIIEVNAIRLSRSALASKAT 996
 QY 894 MQPL-----SLTPOCK-----PEVSANDTSHSTNLN 921
 Db 997 GYPLAIIYSAKIALGYDLISLKNSTIKTTACFEPSSLDYITTKIPRMDLNKFEFASNTMS 1056
 QY 922 SLKLEFENILSLGKNKNIYQELIGQS-----SENFY-----EKIL-----K 957
 Db 1057 SMKSVGEVMSIGRT---FEESI-QKSLRCIDNTYIGFNTYCIDWDEKKIIEELKNPSPK 1112
 QY 958 DSDTFYNESTFNVKSKAD-----INSLNDESKRKKLEE---DINKLX 999
 Db 1113 RIDAIIHQAFHLMMPDKIHELTHIDYWFHFKFYNIYNLQNKLTLEQLSFNDLKYPKK 1172
 QY 1000 -----TLQSFPLYN-----KYKLELRLPDKRTGCKYKMOI 1032
 Db 1173 HGFSDKQIAHLYLSNTSDNNNNNNNISCRTENDVMKREKL-GLPPIKVIDTLSAEF 1231
 QY 1033 KKL------LKEQLESKLNSLNPKHVLONSVEFNKK-----EAEI 1071
 Db 1232 PALTNVLYLTGYQGOEHVPLPLNMKKRKKICTLNNKRNA-----NKKKVHVKNHLYNEV 1283
 QY 1072 AETENT-----LENTKILKHYGLVYKYGESSPLKTLSEESIOTEDN 1115
 Db 1284 VDDKDTOLHKNNNNNNNNNSGVENKCKLNKESYG---YNNSSNCINT---NNINIEEN 1336
 QY 1116 ---YASLENFV-LSKLEGLKIDNLNEKKKLSTLS---SGLHHLIAELKEVI----- 1161
 Db 1337 ICHDITIKNKIKYTIINNSNISNNENE-YNLNCVSRAGSHHYGKEEKSGSDDTNI 1395
 QY 1162 ----KKNKNTGNSPEN--NFDVNNALLESYKKEPLPGTDVATVYSESGLTLBQSQPK 1214
 Db 1396 LSAQNSNNNFSCNNENMKANVDVNLENDTK--REDINTTVEME-GQNSYIINNKE 1452
 QY 1215 PASTHVGAESNTITTSQNVDEVDVITVPIFGSEEDYDGLGOVYVGEAVTPSYIDNIL 1274
 Db 1453 NSSLLKGDEEDIVVNNLKKENNSVI-----NNVDCRRKMDGK---NINDECK 1499
 QY 1275 SKINEYEVLTKPLAGYFSLKOLENNVM-----TFNVNVKDILNSR-----F 1319
 Db 1500 TYKKNKIKDM-----GLNNIYDELSTNGTSHSTNDHLYLDFNTSDEIGNKKMMDMYL 1553
 QY 1320 NKRENFKNVLESDLIPYKDLTSSNVYKDPYKFLNKEKRDKFLSSYNIKDSIDTDIN-- 1377
 Db 1554 SKKESISN-----KNPGNSYYVDSVYNNNEYKIKMKKELIDNENLNDENNNNNVMN 1604
 QY 1378 -----FANDVLGYKKIIEKYSKSDLSIKRY---INDKQENENEKYLPLNNIETLY 1425
 Db 1605 CSNYNNASAFVNGDRDNDENDCIEKNMDHTYKHYNRLNNRSTNERMMLDVNN----- 1659
 QY 1426 KTVNDKIDLEVIHLEAKVANT--YEKSNEVVKIKELNYLKTIDOKLADFKNNNEVGA 1483
 Db 1660 -----EKESNHEKGRNGLKNKKNEKMEKN-KGKNKD-KKNYHYVN-H 1701
 QY 1484 DLSTDYHNNLLTKF 1498
 Db 1702 KRNNEYNSNNIESKF 1716

Search completed: October 27, 2001, 15:52:01
 Job time: 9447 sec

117: gp_est148:*
118: gp_est149:*
119: gp_est150:*
120: gp_est151:*
121: gp_est152:*
122: gp_est153:*
123: gp_est154:*
124: gp_est155:*
125: gp_est156:*
126: gp_est157:*
127: gp_est158:*
128: gp_est159:*
129: gp_est160:*
130: gp_est161:*
131: gp_est162:*
132: gp_est163:*
133: gp_est164:*
134: gp_est165:*
135: gp_est166:*
136: gp_est175:*
137: gp_est176:*
138: gp_est177:*
139: gp_est178:*
140: gp_est179:*
141: gp_est180:*
142: gp_est181:*
143: gp_est182:*
144: gp_est183:*
145: gp_est184:*
146: gp_est185:*
147: gp_est186:*
148: gp_est187:*
149: gp_est188:*
150: gp_est189:*
151: gp_est190:*
152: gp_est199:*
153: gp_est100:*
154: gp_est101:*
155: gp_est102:*
156: gp_est103:*
157: gp_est104:*
158: gp_est105:*
159: gp_est106:*
160: gp_est167:*
161: gp_est168:*
162: gp_est169:*
163: gp_est170:*
164: gp_est171:*
165: gp_est172:*
166: gp_est173:*
167: gp_est174:*
168: gp_est191:*
169: gp_est192:*
170: gp_est193:*
171: gp_est194:*
172: gp_est195:*
173: gp_est196:*
174: gp_est197:*
175: gp_est198:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estp11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_hic:*
187: gp_est107:*
188: gp_est108:*
189: gp_est109:*

190: gp_est110:*
191: gp_est111:*
192: gp_hic:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pin1:*
208: em_gss_pin2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gp_gss1:*
219: gp_gss2:*
220: gp_gss3:*
221: gp_gss4:*
222: gp_gss5:*
223: gp_gss6:*
224: gp_gss7:*
225: gp_gss8:*
226: gp_gss9:*
227: gp_gss10:*
228: gp_gss11:*
229: gp_gss12:*
230: gp_gss13:*
231: gp_gss14:*
232: gp_gss15:*
233: gp_gss16:*
234: gp_gss17:*
235: gp_gss18:*
236: gp_gss19:*
237: gp_gss20:*
238: gp_gss21:*
239: gp_gss22:*
240: gp_gss23:*
241: gp_gss24:*
242: gp_gss25:*
243: gp_gss26:*
244: gp_gss27:*
245: gp_gss28:*
246: gp_gss29:*
247: gp_gss30:*
248: gp_gss31:*
249: gp_gss32:*
250: gp_gss33:*
251: gp_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gp_gss35:*
257: gp_gss36:*
258: gp_gss37:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

FEATURES	source
<p>AUTHORS Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S. TITLE FULL-malaria: a database for a full-length enriched CDNA library from human malaria parasite, Plasmodium falciparum JOURNAL Nucleic Acids Res. 29 (1), 70-71 (2001) MEDLINE 20574754 COMMENT Contact: Junichi Watanabe Institute of Medical Science The University of Tokyo, Department of Parasitology 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan Tel: 81-3-5449-5378 Fax: 81-3-5449-5410 Email: jwatanabe@nagane.ims.u-tokyo.ac.jp Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano,S. S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).</p>	<p>BASE COUNT 229 a 51 c 52 g 146 t 22 others ORIGIN</p>
<p>FEATURES Location/Qualifiers 1..500</p>	<p> /organism="Plasmodium falciparum" /strain="3D7" /db_xref="taxon:5833" /clone="XPn3463" /clone_id="Sugano Malaria cDNA library" /dev_stage="erythrocytic stage" /dev_stage="erythrocytic stage" </p>
<p>Query Match Best Local Similarity 60.2%; Score 168.8; DB 107; Length 500; Matches 266; Conservative 0; Mismatches 176; Indels 0; Gaps 0;</p>	<p> 3.4%; Score 168.8; DB 107; Length 500; Best Local Similarity 60.2%; Pred. No. 6.8e-32; Matches 266; Conservative 0; Mismatches 176; Indels 0; Gaps 0; </p>
<p> QY 2939 atatacaactctcttaacgatgatcttaaacgtaagaagcttggaaagagacatcaataagc 2998 Db 1 ANAGNAATTCACAGTTTGAACCGGAAACAAAAAATGATTAATGATGAAATTAATAAAT 60 OY 2999 tgaagaagacacactgcaactgagctcgagcctgtacaacaagtacaaactgaaactggaga 3058 Db 61 TTAAGGATACCTTACAGTTATTCATTGATTTATATATATATATATATATATATATATAT 120 OY 3059 gactcttcgacaagaagaagacagctcggaagataaagatgcagaatcgaagtctatc 3118 Db 121 GATTATTTAATAGAGAAAAAGACTTGGCCAAAGCAAAATGCAAAATTTAAAAACTTACT 180 OY 3119 tgcctaagaagacgcttgaagaacaactcaactcagcaacaatccgagaacagctactgc 3178 Db 181 TATTAAAGAACAAATTAACATCAAAATTTGAATTCCTTAATTAACCATATATGATTATAC 240 OY 3179 agaactctcgtgtcttccaagaagaagagccgagatcgccgagacgaagaagaaca 3238 Db 241 AAAACTTTTCTGTTTCTTTTAAACANANAAAAAGAGCTGAATATACNCAAACTGANAACA 300 OY 3239 ctctggagaagacccaagattctctccaacaactacaagaagcctctgccaagtataatg 3298 Db 301 CATTTGAAACACANANATATTTATGAAACATTAATAAGCACTTTTAAATTTTAAATG 360 OY 3299 gcgagctctctccctctgaagactctctccgagagagacatccagacggaagtaactcg 3358 Db 361 GTGATTCATCTCCATTTAAACCTTTAAAGTATCAATTCANACAGACATTTATATG 420 OY 3359 ccagcctcgagaactccaagt 3380 Db 421 CCAATTGTAGATTAATTTANAGT 442 </p>	<p> 2939 atatacaactctcttaacgatgatcttaaacgtaagaagcttggaaagagacatcaataagc 2998 1 ANAGNAATTCACAGTTTGAACCGGAAACAAAAAATGATTAATGATGAAATTAATAAAT 60 2999 tgaagaagacacactgcaactgagctcgagcctgtacaacaagtacaaactgaaactggaga 3058 61 TTAAGGATACCTTACAGTTATTCATTGATTTATATATATATATATATATATATATATAT 120 3059 gactcttcgacaagaagaagacagctcggaagataaagatgcagaatcgaagtctatc 3118 121 GATTATTTAATAGAGAAAAAGACTTGGCCAAAGCAAAATGCAAAATTTAAAAACTTACT 180 3119 tgcctaagaagacgcttgaagaacaactcaactcagcaacaatccgagaacagctactgc 3178 181 TATTAAAGAACAAATTAACATCAAAATTTGAATTCCTTAATTAACCATATATGATTATAC 240 3179 agaactctcgtgtcttccaagaagaagagccgagatcgccgagacgaagaagaaca 3238 241 AAAACTTTTCTGTTTCTTTTAAACANANAAAAAGAGCTGAATATACNCAAACTGANAACA 300 3239 ctctggagaagacccaagattctctccaacaactacaagaagcctctgccaagtataatg 3298 301 CATTTGAAACACANANATATTTATGAAACATTAATAAGCACTTTTAAATTTTAAATG 360 3299 gcgagctctctccctctgaagactctctccgagagagacatccagacggaagtaactcg 3358 361 GTGATTCATCTCCATTTAAACCTTTAAAGTATCAATTCANACAGACATTTATATG 420 3359 ccagcctcgagaactccaagt 3380 421 CCAATTGTAGATTAATTTANAGT 442 </p>
<p>RESULT 3 LOCUS A0087573 500 bp mRNA EST 27-JAN-2001 DEFINITION A0087573 Sugano Malaria cDNA library Plasmodium falciparum cDNA clone XPn3232 similar to P.falciparum gp190 (MSA, MSP1, PMMSA) for precursor of major merozoite surface antigens, mRNA sequence. ACCESSION A0087573 VERSION A0087573.1 GI:12389714 FEATURES EST.</p>	<p> 2939 atatacaactctcttaacgatgatcttaaacgtaagaagcttggaaagagacatcaataagc 2998 1 ANAGNAATTCACAGTTTGAACCGGAAACAAAAAATGATTAATGATGAAATTAATAAAT 60 2999 tgaagaagacacactgcaactgagctcgagcctgtacaacaagtacaaactgaaactggaga 3058 61 TTAAGGATACCTTACAGTTATTCATTGATTTATATATATATATATATATATATATATAT 120 3059 gactcttcgacaagaagaagacagctcggaagataaagatgcagaatcgaagtctatc 3118 121 GATTATTTAATAGAGAAAAAGACTTGGCCAAAGCAAAATGCAAAATTTAAAAACTTACT 180 3119 tgcctaagaagacgcttgaagaacaactcaactcagcaacaatccgagaacagctactgc 3178 181 TATTAAAGAACAAATTAACATCAAAATTTGAATTCCTTAATTAACCATATATGATTATAC 240 3179 agaactctcgtgtcttccaagaagaagagccgagatcgccgagacgaagaagaaca 3238 241 AAAACTTTTCTGTTTCTTTTAAACANANAAAAAGAGCTGAATATACNCAAACTGANAACA 300 3239 ctctggagaagacccaagattctctccaacaactacaagaagcctctgccaagtataatg 3298 301 CATTTGAAACACANANATATTTATGAAACATTAATAAGCACTTTTAAATTTTAAATG 360 3299 gcgagctctctccctctgaagactctctccgagagagacatccagacggaagtaactcg 3358 361 GTGATTCATCTCCATTTAAACCTTTAAAGTATCAATTCANACAGACATTTATATG 420 3359 ccagcctcgagaactccaagt 3380 421 CCAATTGTAGATTAATTTANAGT 442 </p>

SOURCE	malaria parasite P. falciparum.
ORGANISM	Plasmodium falciparum
REFERENCE	Eukaryota: Alveolata: Apicomplexa: Haemosporidia; Plasmodium. 1 (bases 1 to 500)
AUTHORS	Matanabe,J., Sasaki,M., Suzuki,Y., and Sugano,S.
TITLE	FULL-malaria: a database for a full-length enriched cDNA library from human malaria parasite, Plasmodium falciparum
JOURNAL	Nucleic Acids Res. 29 (1), 70-71 (2001)
MEDLINE	20574754
COMMENT	Contact: Junichi Matanabe Institute of Medical Science The University of Tokyo, Department of Parasitology 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan Tel.: 81-3-5449-5378 Fax: 81-3-5449-5410 Email: jmatanabem@nagare.ims.u-tokyo.ac.jp Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano, S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).
FEATURES	Location/Qualifiers
Source	1..500 /organism="Plasmodium falciparum" /strain="3D7" /db_xref="taxon:5833" /clone="XPEN5232" /clone_lib="Sugano Malaria cDNA library" /dev_stage="erythrocytic stage"
BASE COUNT	237 a 52 c 58 g 153 t
ORIGIN	
Query Match	3.2%. Score 158; DB 107; Length 500;
Best Local Similarity	59.3%; Pred. No. 3.8e-29;
Matches 288; Conservative	0; Mismatches 195; Indels 3; Gaps 1;
OY 1533	cttcgagaagaatgaaggaaaacggctactcttacaagtgtagaagaacctgaccaccataa 1592
Dd 15	cgtgatataaatattttcaattgcacagtatattacatttatattttgaaaaaacaaagctatatata 74
OY 1593	taccttgcacctcctaatacgaaattcgaacataatcctlagaagctcacaccaagctcttaa 1652
Dd 75	ttaaatTTTCAATCCTGATTAATTTCGTATATAATGTTCAAAAATTTAAAAAAGCGCTCTTC 134
OY 1653	glataatggaactactctctcggaacacttgttctgtagaagaagaactaaglatataa 1712
Dd 135	atattCTGAAGATTAATCTTTAAAGAAAAGAATTTCTGAAAAAGATTTTAATCATTTATTA 194
OY 1713	gaatctcataagtaagatcgaagaacgaagatcgagcgcgttgttggaacattagaagga 1772
Dd 195	TACTTTGAAAACCTGGCCCTCGAAGCTGATATATAAAAAATTAAACGAGAAATTAAGAGTAG 254
OY 1773	tgaagaacagtttgttctgtagaagaagatatacaaagaacgaaataaaccagataagaagat 1832
Dd 255	TGAAAACAATAATTCTAGAAAAAAATTTTAAAGACATAACACATTCAGCAANTG---GTTCC 311
OY 1833	ccctggaagctctcgataltgtttaaagtcacaagtcgagaagtgctcccatgaacaagat 1892
Dd 312	CTTAGAATATCTGATTTGTAATTTGTAATTAACAAGTACAAAAGCTTTTATTAATTTAAAAAAT 371
OY 1893	tgaatgaactcaagaagactcaactcatctctgaagaagcttgagttaaaactaataataca 1952
Dd 372	AGAAACACTTAAGAAAAGATGAGATTAATTTTTTAAAAAATGCACAACTTAAAAAGATGATTCCA 431
OY 1953	tgttcgcaataglttaagaacgagaagataagcgaagaacctactacccttcgtaactaa 2012
Dd 432	TGTACCAATATTTTATAAACCAACAATAATTAACGAGAACCATATTTTAAATTTGTTTAA 491
OY 2013	gaaaga 2018
Dd 492	AAAAGA 497

DB	283	TTTAATTATTGTAATTCAGAACCAAAAAACACACTTTTAAAAAGTTACAAATATATATAA	342
Oy	4117	gactccatcgacaccgatatacoa	4139
Db	343	GAATCAGTAGAAAAATGATATTATA	365
RESULT	6		
LOCUS	AU086246	500 bp	EST
DEFINITION	AU086246 Sugano Malaria CDNA library Plasmodium falciparum CDNA clone XPFn2175 similar to P.falciparum gp190 (MSA1, MSP1, PMMSA)		27-JAN-2001
ACCESSION	AU086246		
VERSION	AU086246.1	GI:12388387	
KEYWORDS	EST.		
SOURCE	malaria parasite P. falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.		
AUTHORS	1 (bases 1 to 500)		
TITLE	Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.		
JOURNAL	FULL-malaria: a database for a full-length enriched CDNA library from human malaria parasite, Plasmodium falciparum		
MEDLINE	Nucleic Acids Res. 29 (1), 70-71 (2001)		
COMMENT	20574754		
	Contact: Junichi Watanabe		
	Institute of Medical Science		
	The University of Tokyo, Department of Parasitology		
	4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan		
	Tel: 81-3-5449-5378		
	Fax: 81-3-5449-5410		
	Email: jwatanabe@med.s.u-tokyo.ac.jp		
	Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano,S.		
	Construction and characterization of a full length-enriched and		
	a 5'-end-enriched CDNA library Gene 200 (1-2), 149-156 (1997).		
FEATURES			
source	location/Qualifiers		
	1..500		
	/organism="Plasmodium falciparum"		
	/strain="3D7"		
	/db_xref="taxon:5833"		
	/clone="XPFn2175"		
	/clone_lib="Sugano Malaria CDNA library"		
	/dev_stage="erythrocytic stage"		
BASE COUNT	244 a 69 c 72 g 115 t		
ORIGIN			
Query Match	2.2%: Score 108.8; DB 107; Length 500;		
Best Local Similarity	58.2%: Pred. No. 1.3e-16;		
Matches 191: Conservative	0: Mismatches 137; Indels 0; Gaps 0;		
Oy	1745	agacgctgtttgagacattgaagaagaatgaagaacgtgttttgagaagaagattaca	1804
Db	19	ATACAAATTAATTACAGAGAATAAAGAGTGTGTAACAAACAAATCTCGAAAAAATTTA	78
Oy	1805	aagcagaataataaaccaagatgagaagatcctggaggtctcgcatctgtttaaagtcgaag	1864
Db	79	AAGGACTACACATCTTACACCAATGGTCTCTTAGAAGATCTGATTTGTAATAATTACAG	138
Oy	1865	tgcagaaggtctcctctctgacaagaatttgtaactaagaagctcaactcattcga	1924
Db	139	TACAAAAAGCTTTTATTATTATAAATAAAGAGACTTAAGAAAGATGACAAATTTATTTAA	198
Oy	1925	agaacgttgagtttaaacataatatacatctgtgcccgaatagttatataagcaggagaatgaac	1984
Db	199	AAAATGCAACACTAATAAGTAGTATTCAATGTATGCCAAATTTATTATAACCAACCAAAATTAAC	258
Oy	1985	aggaacatactactcctcactcgtaactcaagaagaagatgacaaactgaaagtgttcacgc	2044
Db	259	CAGAGCCATATTATTATTATTATTTAAATAAAGAGTGTGATTAATAATTAAGCAATTTATAC	318
Oy	2045	ccaagctcgagagctgatacaagaaga	2072

DB	319	CAAAGCTAAACACATGTTAAGAAGA	346
RESULT	7		
N97689			
LOCUS	N97689	400 bp	mRNA
DEFINITION	1114C3 czapPPDd2.1, Debopam Chakrabarti Plasmodium falciparum cDNA		EST.
ACCESSION	N97689		
VERSION	N97689.1	GI:1674723	
KEYWORDS	EST.		
SOURCE	Malaria parasite P. falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	Dame,J.B., Annot,D.E., Bourke,P., Chakrabarti,D., Christodoulou,Z., 1 (bases 1 to 400) Coppel,R.R., Comnan,A., Craig,A., Fischer,K., Foster,J., Goodman,N., Hinterberg,K., Holder,A.A., Holt,D., Kemp,D., Lanzer,M., Lim,A., Newbold,C., Ravetch,J.V., Reddy,G.R., Rubio,J., Schuster,S.M., Su, X.-Z., Thompson,J.K., Vital,F., Wellens,T.E. and Werner,E. Current status of the Plasmodium falciparum genome project Mol. Biochem. Parasitol. 79, 1-12 (1996)		
TITLE	97001675		
JOURNAL	Contact: Debopam Chakrabarti		
MEDLINE	Department of Molecular Biology and Microbiology		
COMMENT	University of Central Florida		
	Orlando, FL 32816-2360		
	Tel: 407 384 2061		
	Fax: 407 384 3095		
	Email: dcchakepegasus.cc.ucf.edu		
	Seq primer: T3.		
FEATURES	Location/Qualifiers		
source	1..400		
	/organism="Plasmodium falciparum"		
	/strain="Pd2"		
	/db_xref="taxon:5833"		
	/clone="PF1114C"		
	/clone_lib="czapPPDd2.1, Debopam Chakrabarti"		
	/lab_host="E. coli XL-1 blue"		
	/note="Vector: Lambda ZAP II; Site_1: Ecor I; Site_2: Xho		
	I; PolyA+ RNA, from asynchronous blood stage parasites of		
	the Pd2 isolate cultured in vitro, was reverse transcribed		
	using an oligo dt-Xho I primer. Second strand was		
	prepared using RNase H and DNA polymerase I. Ecor I		
	adapters were ligated to the cDNA, and it was digested		
	with Xho I. Prepared fragments were ligated into Ecor I +		
	Xho I digested lambda ZAP II vector."		
BASE COUNT	177 a 58 c 53 g 112 t		
ORIGIN			
Query Match	2.1%; Score 101.6; DB 187; Length 400;		
Best Local Similarity	60.5%; Pred. No. 8,3e-15;		
Matches 167; Conservative	0; Mismatches 109; Indels 0; Gaps 0;		
OY	3316	aagactctctcgcgagagacatccagaccgagatactaaccgacctcgagaacttc	3375
Db	2	AAAATTGAAGGAGTATCAATTCACAACACAGAATATTTATGCCAATTGTAAAAAATTT	61
OY	3376	aaggtccgtctaaagtctgaaggcaagtgtaagacaaccggaacctggagaagaaga	3435
Db	62	AGAGATTTAAGTAAATAATGCATGGAAACTCATATGTAATTTTACATTATTAAGAAAGAAAAA	121
OY	3436	ctcaagctacctctcagcggagctgatcacctgatcgccgagctcaaaggaattaa	3495
Db	122	TTATCTTTCTTATCAAGTGCAATACATCATTTAATTTACTGATTAAGAAGGATATAAAA	181
OY	3496	aacaagaactaacaccgcycaatatagcccagaagaaataataagaacylgaataacgcatcy	3555
Db	182	AATAAATAATTATACAGTAATTTCTCCAGTGGAATAATATTAAGAAGTTAACGAAGCTTTA	241
OY	3556	gaatttacaagaagttcttcgtcctgaagaagacagat	3591

Oy	1055	agacaggagaagatgcgaagagtctgccaaaccatttgctcaaatgacttctcct	1114
Db	785	NNNNAAAAAAAAAANNAAAAANNNNNAANAATAAAATTAATAAAAATTTTAAAAANNN	726
Oy	1115	ttaactgatcccttcgtgagctactctcgtagagaagaataaagaatatagacct	1174
Db	725	NANANAAAAAANNAANNNNANNAATAAAAAATTTTTTANNNNNNNTAAAAANAAAA	666
Oy	1175	cggccaagctcgagacaagaatcaaaccgaaccttaaigtatccccatgltgtgaagt	1234
Db	665	AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNNN	606
Oy	1235	aacctcttgtcataagatatacacaaagctctccaagcgctcaatagcttcgtgagct	1294
Db	605	AAMNNANNN	546
Oy	1295	tgattaacccttcgatlatatcgaaaagaccctctaagaatatctcacagacatgaga	1354
Db	545	NN	486
Oy	1355	gaaagaagtattcaacgaaatcaagagagaatcaaaatltsгаагаагаатгата	1414
Db	485	AAANNN	426
Oy	1415	gtcgacaagaagafttacgaagaccgcccgaacaaagttcaaacgatalcatcactaaagtatg	1474
Db	425	AAAAAAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAAN	366
Oy	1475	aaagctgtcgaaagagatcatatctccaatlcacaaataacatcgacctgaccaact	1534
Db	365	NNNNANANANANANNNANNNANNAANNAANNNNNNNNNNNNNNNNNNNNNNNNNNN	306
Oy	1535	tcgagaaaaatgatyggaaaacytatccttcaaaagtlgsгаааасгыаассааааа	1594
Db	305	NN	246
Oy	1595	ccttgcactctatgagaatcttaagataatcttgagaagccaccacaagccttaagt	1654
Db	245	NN	186
Oy	1655	atatgagagactatctctcctcgagacatgtgtlgysaaagaactaaagtattacaga	1714
Db	185	NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGAAGCNCNNCCNNNNNN	126
Oy	1715	atctcataagtaagatcgaaaacogagatcgagacgc	1750
Db	125	NNNC	90
RESULT_12			
LOCUS			
AZ813205	540 bp	DNA	GSS
DEFINITION	M4008GC24F Mouse 10kb plasmid UUGClm library Mus musculus genomic		20-FEB-2001
ACCESSION	clone UUGC24M080C24 F.	DNA sequence.	
VERSION	AZ813205		
KEYWORDS	AZ813205.1 GI:12983113		
SOURCE	GSS.		
ORGANISM	house mouse. Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus		
REFERENCE	1 (bases 1 to 540)		
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beecorn,T., Duval,B., Hamll,C., Islam,H., Longacre,S., Mahmud,M., Meenen,E., Pedersen,T., Reilly , M., Rose,M., Rose,R., Stokes,R., Tingey.A., von Niederhausern,A. and Wright.D., Weiss.R.		
TITLE	Mouse whole genome scaffolding with paired end reads from 10kbp plasmid inserts		
JOURNAL COMMENT	(Unpublished (2000)) Contact: Robert B. Weiss University of Utah Genome Center		

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
 84112, USA
 Tel.: 801 585 5606
 Fax: 801 585 7177
 Email: ddumgenetics.utah.edu
 Insert Length: 1000 Std Error: 0.00
 Plate: 0080 row: C column: 24
 Seq primer: CGTGTAAACGACGCGCAGT
 Class: plasmid ends
 High quality sequence stop: 540.
 Location/Qualifiers
 1..540

```

BASE COUNT      284 a      54 c      151 g      51 t
ORIGIN
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U0GC2M0080C24"
/clone_id="Mouse 10Kb plasmid U0GCI1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (g114732114[gb|AF129072.1], a copy-number
inducible derivative of plasmid RL. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

```

Query Match	1.5%	Score 73.8	DB 250	Length 540
Best Local Similarity	49.9%	Pred. No. 1	1e-07	
Matches 186	Conservative	0	Mismatches 187	Indels 0
			Gaps	0
Qy 731	atattaaagaataaagaagacccatcgagaacatttaacgagctgatcgaaatccaaa	790		
Db 115	AGATGAGGAAGAAAGAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	174		
Qy 791	agaccatagacataaataaagaatgcacccaagaggaagaagaagaagcttgtaccgg	850		
Db 175	AGAAAGAAAGAAACAACTGTGAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAA	234		
Qy 851	cccagtcgacctgtccatctatatacaaacgcgttgaagaagccataacctcaacgcg	910		
Db 235	AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	294		
Qy 911	tactgtagaagcgcatatgacacccctcaagaagaatgaaatalcataaagaactgtcgaca	970		
Db 295	AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	354		
Qy 971	agatttaagatataagaatctctcgccagcgaacctctgggaacaccccttaacgcgtgc	1030		
Db 355	AGAAAGACTGAGAGAGAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	414		
Qy 1031	tgagcaagaacaagaagatagagagcagcaagaagaagatccccaacaa	1090		
Db 415	AGAACAAAGAAACAAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	474		
Qy 1091	ttaagtccaacat 1103			
Db 475	GCCAGGACCAATAT 487			

RESULT 13
BE420745
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

BE420745 1885 bp mRNA EST 24-JUL-2000
HMM002.B02 ITEC HMM Barley Leaf Library Hordeum vulgare cDNA clone
HMM002.B02, mRNA sequence.
BE420745
BE420745.1 GI:9418588
EST.
barley.
Hordeum vulgare
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
1 (bases 1 to 1885)
Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier
, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,
Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P.,
Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, T.,
Pechioni, N., Quilez, C., Schuch, W., Selvaraj, G., Shariflou, M.,
Sorrells, M., Warburton, M. and Wenzel, G.
International Triticeae EST Cooperative (ITEC): Production of
Expressed Sequence Tags for Species of the Triticeae
Unpublished (2000)
Contact: Herrmann RG
Botanisches Institut der LMU
Menzinger Str. 67, D-80638 Muenchen GERMANY
Fax: 49 30 171683
Email: herrmann@botanik.biologie.uni-muenchen.de
International Triticeae EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.
Location/Qualifiers
1. 1885
/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="taxon:4513"
/clone="HMM002.B02"
/clone_1id="ITEC HMM Barley Leaf Library"
/lissu_type="leaf"
/dev_stage="14 day old"
/note="Vector: pBluescriptSK(-); 850 bp average insert
size."
BASE COUNT 1138 a 219 c 212 g 176 t 140 others
ORIGIN

Query Match 1.48; Score 71.6; DB 167; Length 1885;
Best Local Similarity 37.68; Pred. No. 6.4e-07;
Matches 362; Conservative 0; Mismatches 602; Indels 0; Gaps 0;

Qy 631 aagatcagagcgaacgagcttgacgtatgagaagctgctcgtcgtatcgcgaagct 690
||| |
Db 892 AA 951
||| |
Qy 691 ctgcacacatcaagacaatctggaagaatggaatatataaagaataagaag 750
||| |
Db 952 NNN 1011
||| |
Qy 751 accatcgagaacatcaagcgtcgtcgaagaatccaaagaacatagacaaaataag 810
||| |
Db 1012 AAA 1071
||| |
Qy 811 aatgcacccaagagggaagaagaagaagctgtacccagccagtaagactgtcctc 870
||| |
Db 1072 AAA 1131
||| |
Qy 871 tataacaacagctgaagaagccatacctcagctcaggaagcgtacgtggaagcgtacagc 930
||| |
Db 1132 AAA 1191
||| |
Qy 931 accctcaagaagaatgaatatatacaagaactgctcgacaagatgaataagaat 990
||| |

Db 1192 AAA 1251
Qy 991 cctccgccaagcactcgggaacacccctaacagcgtcgtggaagaagaagata 1050
||| |
Db 1252 AAA 1311
||| |
Qy 1051 gagagcaggaagaagatcaagaagatcgccaaacattgaatltcaatagattct 1110
||| |
Db 1312 AAA 1371
||| |
Qy 1111 cctcttactgacctcgtgagctgagctactctgaggaagaagaataagatagac 1170
||| |
Db 1372 AAA 1431
||| |
Qy 1171 atccgcccaagctcgagacaagaatcaacgcgaacctaatgatatcccaatggtg 1230
||| |
Db 1432 AAA 1491
||| |
Qy 1231 acgtaccctctgtcttatacagatatacacaacgctctcaacgagctcattagctcgt 1290
||| |
Db 1492 AAA 1551
||| |
Qy 1291 gactgattaacctctcgtatatacgaagaagacctcttaagaatatctacagacat 1350
||| |
Db 1552 AAA 1611
||| |
Qy 1351 gagagaagaagtttatcaacgaatcaaggaagaatcgaaatggaagaagaatt 1410
||| |
Db 1612 AAA 1671
||| |
Qy 1411 gagatgcaagaagaatgtaacgaagccgcgaagaagttcaacgatatcaactaagag 1470
||| |
Db 1672 AAA 1731
||| |
Qy 1471 tatgaagaagcgtcgaagcgtatcgtatccaaatcaacataacatgactgacc 1530
||| |
Db 1732 NAA 1791
||| |
Qy 1531 aactcgaagaatgatacggaagaacgctaccttcaagaagtggagaacgtgacacacat 1590
||| |
Db 1792 AAA 1851
||| |
Qy 1591 aata 1594
||| |
Db 1852 AAAA 1855

RESULT 14
A2548467
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

A2548467 908 bp DNA GSS 14-NOV-2000
ENTER30TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
A2548467
A2548467.1 GI:11172102
GSS.
Entamoeba histolytica.
Entamoeba histolytica
Eukaryota: Entamoebidae; Entamoeba.
1 (bases 1 to 908)
Loftus, B., Van Aken, S. and Fraser, C.
Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse

Query Match 1.4%; Score 68.4; DB 244; Length 596;
 Best Local Similarity 50.3%; Pred. No. 2.8e-06;
 Matches 168; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

```

QY 736 aaaaagaataagaagaccatcgagaacatlaacgagctgtatcgagaatcccaaaagacc 795
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 29 AAGAACCAAGAAAGAACACAGACAGAACAAAGAAAGAAAGAACAAAGAACAGAACAG 88
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 796 atagacaaaataagaatgcacccaaggaagaaagaaagaaagtgtaccagccag 855
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 89 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAACAAA 148
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 856 tacgacctgtccatctataacaacagcttgaagaagcccaataacctcatcagctactg 915
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 149 AAGAACCAAGAAAGAACACAGACAGAACAAAGAAAGAAAGAACAAAGAACAGAACAG 208
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 916 gagaagcgcatagcacccctcaagaagaatlgaaatatcaagaactgtctgcacaagatt 975
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 209 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 268
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 976 aatgaattaaagaatctctccgcagcccaactctggaacacccctaaacgctgtgac 1035
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 269 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 328
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1036 aagacaagaagaatagaggagcagcagaagaaga 1069
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 329 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 362
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    
```

Search completed: October 27, 2001, 10:44:19
 Job time: 5135 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2001, 13:13:29 ; Search time 34.9 Seconds
(without alignments)
2847.069 Million cell updates/sec

Title: US-09-269-874-3

Perfect score: 8424
Sequence: 1 MKIIFLCFLFIINTQCV.....SNFLGSLFLMLILYSPI 1639

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601:*

- 1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8418	99.9	1639	19	AAW54145
2	8139.5	96.6	1634	6	AAW50777
3	1881	22.3	355	20	AAW09372
4	1881	22.3	355	20	AAW05832
5	1881	22.3	361	20	AAW09373
6	1881	22.3	361	20	AAW05833
7	1869	22.2	376	20	AAW09374
8	1869	22.2	376	20	AAW05834
9	1333	15.8	262	22	AAW37610
10	913	10.8	355	9	AAW80547
11	910.5	10.8	219	9	AAW82523

12	905	10.7	219	9	AAW80544	Polypeptide p190-1
13	794	9.4	244	6	AAW50304	Storage-specific,
14	677	8.0	151	9	AAW80546	Polypeptide p190-2
15	637	7.6	116	18	AAW36103	PfMSP1(p19)A prote
16	637	7.6	116	18	AAW22592	PfMSP1(p19)A prote
17	560	6.6	108	22	AAW37609	Merozoite surface
18	559.5	6.6	1979	21	AAW18171	Plasmodium falcipa
19	556	6.6	96	22	AAW37608	Merozoite surface
20	539	6.4	127	18	AAW22593	PfMSP1(p19)S prote
21	539	6.4	127	18	AAW36102	PfMSP1(p19)S prote
22	455.5	5.4	1254	11	AAW07503	Merozoite apical-en
23	455.5	5.4	1254	18	AAW24575	Merozoite apical-e
24	453.5	5.4	2485	21	AAW18172	Plasmodium falcipa
25	442.5	5.3	3973	21	AAW18253	Plasmodium falcipa
26	431	5.1	980	21	AAW18254	Plasmodium falcipa
27	424	5.0	102	6	AAW50303	Storage-specific,
28	414	4.9	2010	21	AAW18218	Plasmodium falcipa
29	411.5	4.9	1516	21	AAW18195	Plasmodium falcipa
30	408.5	4.8	2539	21	AAW18198	Plasmodium falcipa
31	399	4.7	2013	21	AAW18265	Plasmodium falcipa
32	393	4.7	2954	20	AAW01632	Amino acid sequenc
33	388.5	4.6	2295	21	AAW18180	Plasmodium falcipa
34	387.5	4.6	4134	20	AAW31946	Plasmodium falcipa
35	383.5	4.6	1308	21	AAW18167	Plasmodium falcipa
36	383.5	4.6	2573	21	AAW18234	Plasmodium falcipa
37	379.5	4.5	2206	21	AAW18254	Plasmodium falcipa
38	372	4.4	2500	21	AAW18272	Plasmodium falcipa
39	370	4.4	1411	17	AAW02258	Nucleolar/endosoma
40	369	4.4	1087	20	AAW19935	B. burgdorferi ant
41	369	4.4	1119	20	AAW19934	B. burgdorferi ant
42	365.5	4.3	2380	21	AAW18315	Plasmodium falcipa
43	362.5	4.3	1712	21	AAW18295	Plasmodium falcipa
44	360.5	4.3	1714	21	AAW18275	Plasmodium falcipa
45	357.5	4.2	1817	21	AAW18255	Plasmodium falcipa

ALIGNMENTS

RESULT 1	
AAW54145	AAW54145 standard; Protein: 1639 AA.
XX	
AC	AAW54145;
XX	
DT	23-SEP-1998 (first entry)
XX	
DE	P. falciparum synthetic gp190 protein.
XX	
KW	gp190: malaria; MSP-1; merozoite surface protein; stability; vaccine;
KM	monoclonal antibody; passive immunisation; parasite.
XX	
OS	Plasmodium falciparum.
XX	
PN	Synthetic.
XX	
PD	W09814583-A2.
XX	
PD	09-APR-1998.
XX	
PF	02-OCT-1997; 97W0-EP05441.
XX	
PR	02-OCT-1996; 96DE-4040817.
XX	
PA	(BUJA/) BUJAR H.
XX	
PI	Bujard H, Pan W, Tolle R;
XX	
XX	
DR	WPI: 1998-240088/21.
XX	
DR	N-PSDB: AAV21451, AAV35363.
XX	
PT	Recombinant production of complete gp190/MSP-1 Plasmodium surface
XX	protein - useful in anti-malaria vaccines, also stabilising genes by
PT	reducing their AT content

XX Example 1; Fig 3c; 48pp; German.

CC This sequence represents a modified Plasmodium falciparum gp190/NSP-1
CC (merozoite surface) protein. The gene encoding this protein has been
CC stabilised by reducing the AT content of the nucleotide sequence. Such a
CC protein is useful in vaccines against malaria or for producing monoclonal
CC antibodies (for passive immunisation). The complete gp190 protein can now
CC be produced outside the parasite and has, at least over extended regions,
CC the native pattern of folding. Larger amounts of the protein can be
CC produced recombinantly than would be possible using the parasites as
CC source.

CC Sequence 1639 AA:

Query Match 99.9%; Score 8418; DB 19; Length 1639;
Best Local Similarly 99.9%; Pred. No. 0;
Matches 1638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKIIFFLCSFLFIINQCVTHESEYOELVKRLALEDAVLGYSILFOEKWVLENGTSGT 60
DB 1 MKIIFFLCSFLFIINQCVTHESEYOELVKRLALEDAVLGYSILFOEKWVLENGTSGT 60
QY 61 AVTTSTPGSKGVSAGSGSVASGVSAGSVASGVSAGSGSVASGSGSRRTPSPDSSDS 120
DB 61 avttstpgskgvsagsgsvasgsvasgsvasgsvasgsgsrrtpspdsdsds 120
QY 121 DAKSYADLKHVRNYLTITKELKPOLFDLTNHLITLCDNTHGKRYLDIGEETINELLYK 180
DB 121 daksyadlkhvrnyltitkelpolfdltnhmltldcnthgikryldigyeetinellyk 180
QY 181 LNFEDFLRAKLVNDCANDVCOIPEFNLIKIRANELDVLKTLFVGRKPLDNTKDNVGMED 240
DB 181 lnfedflraklvnvcandvcoipefnlikiraneldvltkflvgrkpldntkdnvgmmed 240
QY 241 YIKKNKKTINENINLEESKKTIDKNKNATKEEKKKLYQAOYDLSYKNOLEBAHLIS 300
DB 241 yikknkktieneninleeskktdkknatkkeeekkklyqaydlsyknqleebahlis 300
QY 301 VLEKRITLTKKNENIKELDOKINIKNPPRNSGNTPTLTLDKKKIIEHKEKEIETAKT 360
DB 301 vlekritletknenikelokiniknpprnsngntptltldkknkieehkekeietakt 360
QY 361 IKFNIIDSLFTDPLELEYLRKKNKNIDISAKVETKESTPEPNYNGVYTPSYNDINNAL 420
DB 361 ikfnidslftdpleleylrkknknidisaavetkestpepnnyngvytpsyndinanal 420
QY 421 NELNSFGDLINPDYTKEPSKNITDNERKKFINEIKKIEKKIESDKKSYEDRSKS 480
DB 421 nelnsfgdlinpdytkepsknitydnerkkfineikkiekkiesdkksyedrsks 480
QY 481 LNDITREYKELINETYSKFNNNIDLTNEKMGKRSYKVEKTLHNHTFASYSNSKHNL 540
DB 481 lnditreyekelinety skfnnnidltnekmgrsykvektlhnhtfasyensknhl 540
QY 541 EKLTKALKYMEDYSILRNIVVEKELKYKNLISKIENEIETLVENIRKKDEQLFEKKTIKD 600
DB 541 ekltkalkymedysilrnivvekelkyknliskieneietlvnenirkkdeqlfekktikd 600
QY 601 ENKPEDEKILEVSDIVKVOVKVLMANKIDELKKTQLILKNVELKHNTHVPSYKOENKOE 660
DB 601 enkpedekilevsvdivkvqvlmankidelkktqlilknvelkhnthvpsykenoe 660
QY 661 PYYLIVLTKKEIDKIKVMPKVESLINEEKKIKTEGOSDNSEPTGEETIGQATTKPGQQ 720
DB 661 pyylivltkkeidkikvmpkveslineekniktegsdnpseptegeitqgatkpgqg 720
QY 721 AGSALIEDSDVOAQOEOKQAPVPVPEAKAQPPTPAPVNNKTEVNSLTDLEKLYE 780
DB 721 agsaledsdvoaqoekqapvpvppeakagpvtppavnnktenvnsldtleklye 780
QY 781 PLNTSYICHKIYILVSHSTMEKILKQYKTKTEESKLSLSCDPLDLFNIQNNIPVMSMF 840

DB 781 flntsyichkiylvshstmeekilkqytkteeskslsscdpldlfniqnpvmsmf 840
QY 841 DSLNNISQLEMEIYKEKEMVNCNLKIKDNOKIKNLLEPAKKVSVTSLSSSQMPLST 900
DB 841 dslnnisqlmeiyekeemvncnlkikdnoklnlleakkvsvtsslsssqmplist 900
QY 901 PQDPEVSANDTSHSNLNSLKLFEINISLGNKNKNIYQCLIOKSSSENYEKILKDS 960
DB 901 pqdpesandtshsnlnslklfeinislgnknkniyqclioqsssenyekilkdsd 960
QY 961 TFWNESTFWKSKADDINSINDESKRRKLEEDINKLKTQLSLFDLYNKYKLEERLPD 1020
DB 961 tfwnefstfwkskaddinsindeskrrkleedinklktqlslfdlynykyleerlpd 1020
QY 1021 KKKTVGKXKMOIKKTLTLKEDESKLNSLNPKHYLVNFSFVFNKKKAEALAEENTLEN 1080
DB 1021 kkktvgkxkmoikkltlkeedesklnslnpkhylvnfsfvfnkkkaealaeentlen 1080
QY 1081 TKILKHYKGLVKKYNGESSPLKTLSESIOTEDNVASLENFKVLSKLECKLKDNLEK 1140
DB 1081 tkilkykglvkkyngeessplktlseesiotednvasklvlsklecklkdnleek 1140
QY 1141 KRLSYLSSGLHHLJAELEKVIKKNKNTGNSPSENNVDVNNALLESYKRFLEPGTDVATVS 1200
DB 1141 krlsyylssglhhljaelekvikknkntgnspsennvdvnnallesykrflepdtvatvs 1200
QY 1201 ESGSDTLEQSGPKRPASTHVAESNTTTSQNVDDVDVIVIPFESSEDYDLDLQOV 1260
DB 1201 esgsdtleqsgpkrpasthvaeesntttsqnvddvdviviipfesedydldlqv 1260
QY 1261 TGEAVTPSVIDINILSKITENEYEVLYKPLAGVRSLSKKOLENNVMTFNVNKKDLNSRFN 1320
DB 1261 tgeavtpsvidinilskiteneyevlykplagvrslskkolennvmtfnvnkkdlnsrfn 1320
QY 1321 KRENKVNLESDDLIPYDLTSSNVTVDPKYFLNKKERDKFLSSYNYIKOSIDPDINPAN 1380
DB 1321 krenkvnlesddlipydltsnvtvdpyflnkkerrdkflssynyikosidpdinpan 1380
QY 1381 DVLGYKILSEKYSKSDSITIKYINDKOGEMEKYLPFLNNIETLYKTVNOKIDLFVHLE 1440
DB 1381 dvlgykilsekysksdstitkyindkogemekylpflnnietlyktnvokidlfvhle 1440
QY 1441 AKVNLTYEKSNSVEVRIKELNYLKTIDOKLADFKKNNNVEGIADLSTDYNNHNLTKFELS 1500
DB 1441 akvnltyeksnsvevrikelnlylktidokladfkknnnvegiadlstdynnhnltkfels 1500
QY 1501 TGMVFNLAKTIVLSNLDGNIOGMLNISQHOVCVKKQCPONSQCFRHLDERECKCLLNK 1560
DB 1501 tgmvfnlaktivlsnldgnioqmlnisqhovcvkkqcpnsgcftrhldereckcllnk 1560
QY 1561 QEGDKVCVENPPTCENNNNGGDADAKCTEEDSGSNGKRTCECRPPSYLPFDGIFPSSS 1620
DB 1561 qegdkvcvenpptcennggdadackteedsgsngkrtcecrppsylpfdgiffcss 1620
QY 1621 NPLGISFLILMLILYSFI 1639
DB 1621 nplgiffilmlillysfi 1639

RESULT 2
AAP50777
ID AAP50777 standard; Protein: 1654 AA.
XX AAP50777:
XX AC AAP50777:
XX DT 30-Sep-1991 (first entry)
XX DE Sequence of the P195 protein of Plasmodium falciparum.
XX KM Malaria vaccine: epitope; antigen: immunogen.
XX OS Plasmodium falciparum.

XX EP154454-A.
XX
XX 11-SEP-1985.
XX
XX 21-FEB-1985; 85EP-0301173.
XX
XX 26-SEP-1984; 84GB-0024340.
XX 22-FEB-1984; 84GB-0004692.
XX 21-FEB-1985; 85GB-0004429.
XX
XX (WELL) WELLCOME FOUNDATION LTD.
XX
XX Holder A. Sandhu J, Odink K, Lockyer M, Riveros-Moreno V;
XX WPI: 1985-224845/37.
XX N-PSDB: AAN50530.
XX
XX Cloned DNA sequence encoding plasmodium falciparum protein -
XX useful for expressing the protein for use in vaccines against
XX malaria
XX
XX Claim 6; Fig 1; 51pp; English.
XX
XX The sequence encoding the P195 protein of Plasmodium falciparum
XX (AAN50530) and a peptide comprising at least one of its epitopes
XX (see AAP50777) are claimed. Also claimed is a vaccine for inducing
XX immunity to malaria comprising the novel peptide or P195 or a
XX peptide comprising at least one epitope when derived from the new
XX DNA sequence, together with a carrier.
XX
XX Sequence 1654 AA:

Query Match 96.6%; Score 8139.5; DB 6; Length 1654;
Best Local Similarity 96.7%; Pred. No. 0;
Matches 1599; Conservative 9; Mismatches 31; Indels 15; Gaps 2;

QY 1 MKIIFFICSPFIITINOCYTHESYOELVKKLEALDEAVLTGYSLFOKEKAVLNEGTSGT 60
DB 1 MKIIFFICSLIFILINQCVCVHESYQELVKKLEALDEAVLTGYSLFQKENVLNEGTSQT 60
QY 61 AVTTSTFGSKSVASGSGSVASGSGSVASGSGSVASGSGSGNSGRTNPSDSDS 120
DB 61 avtstfsgsksvasgsgsvasgsgsvasgsgsvasgsgsgnsrtnpdsds 120
QY 121 DAKSYADLKHVRNYLLTTELKYPQLFDLTNHLTLCDNIGFRKYLIDGYEELNELLK 180
DB 121 daksyadlkhvrnyllttelkypqlfdltnhmltldcnihgfrkylldgyeelnellk 180
QY 181 LNFFEDLLRAKLVNVCANDYQIFPNLIRANELDVLKLVFGYRKLPLDNTKDNVGMED 240
DB 181 lnffedllraklvnvcandyqifpnliraneldvlklvfgyrklpldntkdnvmed 240
QY 241 YIKNKKTIEINILIEESKKTIDKNKNATKEEEKKKLYOQVLSYTKOLEFAHNLS 300
DB 241 yiknkktieinilieeskktdknknateekkklyoqvlstytkolefahnlis 300
QY 241 YIKNKKTIEINILIEESKKTIDKNKNATKEEEKKKLYOQVLSYTKOLEFAHNLS 300
DB 241 yiknkktieinilieeskktdknknateekkklyoqvlstytkolefahnlis 300
QY 301 VLEKRIIDLTKNENIKELDKINEIKNPPRANGTPTLTDKNKKIEEHKEIKETAKT 360
DB 301 vlekridltknenikeldkineiknpprangsptltldknkkieehkeiketakt 360
QY 361 IKFVDSLFDLPELEYLREKKNKIDISAKVETKESTEPNEYNGVYPLSYNDINNAL 420
DB 361 ikfvdsldfdeleylrekknkidisakvetkestepneyngvyplsyndinnal 420
QY 421 NELUSFDLINPFTYKEPSKNITDNERKKFIEIKETIKETIKESDCKSVEDSKS 480
DB 421 nelusfdlinpftykepsknitdnerkkfieiketikeetikesdcksvedsks 480
QY 481 LNDITKEYEKLINIEYDSKFNNDLITNFEKMGKRYSYKVEKELTHNTFASYSKHNL 540
DB 481 lnditkeyeklineiydskfnndlitnfekmgrysyvekelthntfasyskhn 540

QY 541 EKLTKALYMEDYSLRNTVEKEELKYKKNLSKTIENEIETLVENIKKDEOLFEEKITKD 600
DB 541 ekltkalymedyslrntvekeelkykknlsktieneietlvnikkdeolfekkiktd 600
QY 601 ENKPEDEKILEVSDIVKVOVKVLLMNKIDELKKTOLILKYNELKHNIHVPSYQENKOE 660
DB 601 enkpedekilevsdivkvovkvllmnkidelkktolilkynelekhnihvpsyqenkoe 660
QY 661 PYYLIVLKEIDKLKVPKVESLNEEKKNIKTEGOSDNSEPTGEITGOATTKPGQO 720
DB 661 pyylivlkeidklkvpkveslneekkniktegosdnseptgeitgoattpgqo 720
QY 721 AGSALLEGDSVQAQOEQAOPPVYPPVPEAKAQPVPAPVNNKTEVNSLDVLEKYE 780
DB 721 agsallegdsvgaqaeqapppvppvpeakaqpvpapvnnktevnsldvlekye 780
QY 781 FLNNTSYICHKYLIVSHSTPMNEKIKQYKTKKEESKILSSCDPLDLFNQNNIFVMSMF 840
DB 781 flnntsyichkylivshstpmnekikqyktkeeskilsscdpldlfnqnnifvmsmf 840
QY 841 DSLN-----NSLSQLEMEIYEREMVONLYKLNKDNKIKNLEBAKAVTSV 886
DB 841 dslnhvnylwyekelgyvfllimeyekemvonlyklndnkiknlleakavtsv 900
QY 887 KTLSSSSMQPLSTRPOKPEVSAANDPTSHSTNLNLSKLFENIISLCKNNKIYQELIQK 946
DB 887 ktlssssmqplstrpokpevsandptshstnlslsklfeniiislcnnkiyqelirqk 946
QY 947 SSENPEKILKSDPTFYNESFTNFKSKADINSINDESKRKLKEEDNKKTKTLQLSFD 1006
DB 947 ssenpekilkdsptfyneftnfkskaddinsindekkrklkeednkktktlqlsf 1020
QY 1007 LYNNKYKLEERLEPKKTVGYYKMOIKKLLLEQLSEKLSLNNPKHVILONFSVFENK 1066
DB 1007 lynnkykleerlepkktvgyykmoikkllleqlseklslnnpkhvilonfsvfenk 1066
QY 1021 LYNKYKILKLERLEPKKTVGYYKMOIKKLLLEQLSEKLSLNNPKHVILONFSVFENK 1080
DB 1021 lynykylklerlepkktvgyykmoikkllleqlseklslnnpkhvilonfsvf 1080
QY 1067 KEAEIAEENTLENTKILKRYKGLVRYKYNSESSPKTLSESIOTEDNVASLENFKYLS 1126
DB 1067 keaeiaeentlentkilkrykglvrykynseesspktlseesiotednvassenfkyls 1140
QY 1127 KLECKLKDNLMLKREKKSLSYSSGLHHLAELEKVIKKNYNGNSPSENNTPVNNALSEYK 1186
DB 1141 klegkikhnllekkslsysrglhlhlaelkevikknnyngnspsyntvnnalaseyk 1200
QY 1187 KFLPEGIDVATVSESGSDTLEQSQPKKPASTHWGAESNTTTTSQNVDEVDVYIPIF 1246
DB 1201 kflpegidvatvsegsdtleqsqppkpasthwgaesntttsqnvdevdvyilif 1260
QY 1247 GESEEDYDDLQVYVYGEAVPVSVIDNLSKTIENEYEVLYLKLPLAGVRSLSKOLENNVMT 1306
DB 1261 geseedyddlqvvyvgeavpvsvidnlsktieneevlylklplagvrslyskolennvmt 1320
QY 1307 FNVNWKDILNRFNKKRENFKNVLESDLIPYKDLTSSNVVVDYKPLNKKERDRLSSYN 1366
DB 1321 fnvnwkdlisrnfknrenfnvlesdlipykdltsnvvvdykplnkkerdrflssyn 1380
QY 1367 YIKSDIPTDINFANDVLYGYVILSEKYSKSDLSIKKIYNDKOGGENEYKLPPLNIIETLYK 1426
DB 1381 yiksdipdinfandvlygyvilsekyksdlsikkiyndkoggeneyplplniietlyk 1440
QY 1427 TVNDKIDLFTVHLEAKVINTYREKSNVEVKIKELNLYLKTODKLADPKKNNVGIADLS 1486
DB 1441 tvndkidlftvhleakvintyreksnvevkiikelnlylktodkladfknnvfiadls 1500
QY 1487 TDVNHNNLITFKELSTGWFENLAKTVLSNLLDGM-L-OGMLNISOHQCVKQCPNNSGFR 1545
DB 1501 tdyvnhnnlftkfstgwfenlaksvlsnlldwklatyvknftftpmrkklmigsqsgcfr 1560
QY 1546 HLDREDECKLLNKKQSGDKCVENPPTCNENNCGCADAACTEEDSGSNGKKTTCCTCT 1605
DB 1561 hldreedeckllnyksgsksvensnptcnenngcadacteedsgsngkktctcctk 1620

```

QY 1606 PDSPYLFDFGICSSSNFLGISFLILMLILTYSF 1639
Db 1621 pdcyplasmvifcscsnflgisflilmlilysf 1654

RESULT 3
ID AAY09372 standard; Protein: 355 AA.
AC AAY09372:
XX
XX
XX 31-AUG-1999 (first entry)
DE Merozoite surface protein MSP-1-42.
XX
XX MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
XX transgenic animal.
XX Plasmodium falciparum.
XX
XX WO9920774-A2.
XX
XX 29-APR-1999.
XX
XX 20-OCT-1998; 98WO-US22226.
XX
XX 15-MAY-1998; 98US-0085649.
XX 20-OCT-1997; 97US-0062592.
XX
XX (GENZ ) GENZYME TRANSGENICS CORP.
XX
XX Chen LH, Meade H;
XX
XX WPI: 1999-288313/24.
XX
XX P-PSDB; AAX56008.
XX
XX Modified malarial protein for use in anti-malarial vaccines
XX
XX Example; Fig 1; 35pp; English.
XX
XX The present sequence represents a 42 kDa C-terminal portion of
XX malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
XX important target for the development of a vaccine against
XX Plasmodium falciparum. The nucleic acid sequence encoding MSP-1-42
XX has been modified (see AAX56008) compared to the native sequence (see
XX AAX56009) such that 306 nucleotide positions have been replaced to
XX lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
XX instability motifs while maintaining the same protein amino acid
XX sequence. These alterations allow MSP-1-42 to be expressed in
XX mammalian cell culture and in transgenic mice. Native MSP-1-12
XX is known to be difficult to express in cell culture systems. The
XX mammalian cell culture systems or in transgenic animals. The
XX invention allows expression of MSP-1 protein in the milk of
XX transgenic animals, and also provides a DNA vaccine comprising a
XX vector containing the altered MSP-1-42 sequence.
XX
XX Sequence 355 AA:

Query Match 22.3%; Score 1881; DB 20; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.8e-75;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1264 AWPFSVLDNITSKENYEVLKPLAGVHSLSKKOLENNVMTNNVVKDILNSRPNKRE 1323
Db 1 avpfsvidnitskeneyevlylplagvyslkkqlennvmtfnnvkdilnsrfnkre 60

QY 1324 NEKVVLSDDLIPYDLNASSNVVADPKYFLNKKERKDFLSYNYIKDISDIDINPANDVL 1383
Db 61 nfkvvlsddlipydlnassnvvadpkylfnkkrdfllsyynyikdsiddinfandvl 120

QY 1384 GYKYLSEKYSKSLDSIKKYINDKGENEKYLPPLNNIETLYKTVNDKIDLFVHLLEAKV 1443

```

```

Db 121 gyyklisekysdsldslkkyindkgeneky lplfnnietlyektvndkidlfvhlleakv 180
QY 1444 LNTYKESNVEVKIKELNLYKTIDODKADFFKNNNNGVIADLSDYNNHNLITFLSTGM 1503
Db 181 lntykesnvevkiikelnylktidqkladfkknntfviadlstcyhnnlilkflstgm 240
QY 1504 VFENLAKTVLSNLDGNLQGLNLSOHOCVKKOCPONSGCRHLDERECEKCLNLYKQEG 1563
Db 241 vfenlaktvlsnldgnlqglnlsqhqcvkkqcpnsgcfrhldereckclnlykqeg 300
QY 1564 DKCVENPPTCNENNGCDADAKCTEEDSGSGNKKIRICECTKPPSYPLFDGIFCS 1618
Db 301 dkcvenpntcnennngcdadakcteedsngskkircectkppsyplfdgifs 355

RESULT 4
ID AAY05832 standard; Protein: 355 AA.
AC AAY05832:
XX
XX
XX 02-AUG-1999 (first entry)
DE Merozoite surface protein MSP-1-42.
XX
XX MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
XX transgenic animal.
XX Plasmodium falciparum.
XX
XX WO9920766-A2.
XX
XX 29-APR-1999.
XX
XX 20-OCT-1998; 98WO-US22225.
XX
XX 15-MAY-1998; 98US-0085649.
XX 20-OCT-1997; 97US-0062592.
XX
XX (GENZ ) GENZYME TRANSGENICS CORP.
XX
XX Chen LH, Meade H;
XX
XX WPI: 1999-302742/25.
XX
XX N-PSDB; AAX25586.
XX
XX New modified recombinant nucleic acid sequences useful for producing
XX malarial DNA vaccine
XX
XX Disclosure; Fig 1; 43pp; English.
XX
XX The present sequence represents a 42 kDa C-terminal portion of
XX malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
XX important target for the development of a vaccine against
XX Plasmodium falciparum. The nucleic acid sequence encoding MSP-1-42
XX has been modified (see AAX25586) compared to the native sequence (see
XX AAX25587) such that 306 nucleotide positions have been replaced to
XX lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
XX instability motifs while maintaining the same protein amino acid
XX sequence. These alterations allow MSP-1-42 to be expressed in
XX mammalian cell culture and in transgenic mice. The invention
XX provides modified recombinant nucleic acid sequences and methods for
XX increasing the mRNA levels and protein expression of proteins that
XX are difficult to express in cell culture systems, mammalian cell
XX culture systems or in transgenic animals. The preferred difficult
XX protein candidates for expression are those derived from lower
XX organisms such as parasites, bacteria and viruses that have DNA
XX coding sequences of high AT content or which have mRNA instability
XX motifs or rare codons relative to the recombinant expression system
XX to be used. The invention allows expression of MSP-1 protein in
XX the milk of transgenic animals, and also provides a DNA vaccine

```


CC comprising a vector containing the altered MSP-1-42 sequence.
XX Sequence 355 AA;

Query Match 22.3%; Score 1881; DB 20; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.8e-75;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1264 AVTPSVIDNLSKTEIENEYVLYLKPLAGVYRSLSKKOLENNVTFENVVKDILNSRFNKRRE 1323
DB 1 avtpsvidnlsktenienevlylklplagvyrsllkqlennvmtlnvkvkdlinsrfnkre 60
QY 1324 NFKNVLESDDLIPYKDLTSSNVVVKDPYKFLNKKERKDFLSSYNTIKDSIDTDINPANDVL 1383
DB 61 nfkvnlesddlipykdltsnvvvkdpykflnkkrdfklsyntykdsidtdinfandvl 120
QY 1384 GYKILSEKYSKSDLSIKKYYINDKOGENEKYLPLNNIETLYKTVNDKIDLFVTHLEAKV 1443
DB 121 gyykilseskysdlsikkyindkogenekylplnnietlykcvndkidlftvthleakv 180
QY 1444 LNTYKESNVEVKIKELNYLKTIDOKLADFRKNNNFVGIADLSTDYHNNLLTFELSTGM 1503
DB 181 lntykesnvevkiikelnylktldokladfrknnnfvgiadlstdyhnnlltfelstgm 240
QY 1504 VFENLAKTVLSNLDGMLQGLMNTISOHCYVKKCPONSQCFRHLDERECCCLNTYKOE 1563
DB 241 vfenlaktvlsnldgmlnsgmlnsqhcvcvkcpqnsqcftrhldereccclntykoe 300
QY 1564 DKCVENPPTCENNNGCGDADAKCTEEDSGSNGKRITCECTKPSYPLFDGIFCS 1618
DB 301 dkcvenpntcennngcgdadakcteedsngsklitcectkpsypfldgifcs 355

RESULT 5
ID AAY09373 standard; Protein: 361 AA.

XX AAY09373;

DT 31-AUG-1999 (first entry)

XX Merozoite surface protein MSP-1-42.

XX MSP-1; merozoite surface protein; malaria; vaccine;
KM protein engineering; protein expression; codon usage;
transgenic animal.

XX Plasmodium falciparum.

XX WO9920774-A2.

XX 29-APR-1999.

XX 20-OCT-1998; 98WO-US22226.

XX 15-MAY-1998; 98US-0085649.

XX 20-OCT-1997; 97US-0062592.

XX (GEN2) GENZYME TRANSGENICS CORP.

XX Chen LH, Meade H;

XX WPI; 1999-288313/24.

XX P-PsDB; AAX56009.

XX Modified malarial protein for use in anti-malarial vaccines

XX Example; Fig 2; 35pp; English.

XX This present sequence comprises a 42 kDa C-terminal portion of
CC malarial merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against

CC Plasmodium falciparum. The C-terminal end of the sequence is
CC modified to include a 6xHis tag. A nucleic acid (see AAX56008)
CC encoding MSP-1-42 has been modified according to a method
CC of the invention in order to improve expression in mammalian cells
CC and in transgenic animals by reducing the AT content and removing
CC mRNA instability motifs. The invention allows expression of
CC MSP-1-42 in the milk of transgenic animals, and also provides a DNA
CC vaccine comprising a vector containing the altered MSP-1-42 nucleic
CC acid.

XX Sequence 361 AA;

Query Match 22.3%; Score 1881; DB 20; Length 361;
Best Local Similarity 100.0%; Pred. No. 1.9e-75;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1264 AVTPSVIDNLSKTEIENEYVLYLKPLAGVYRSLSKKOLENNVTFENVVKDILNSRFNKRRE 1323
DB 1 avtpsvidnlsktenienevlylklplagvyrsllkqlennvmtlnvkvkdlinsrfnkre 60
QY 1324 NFKNVLESDDLIPYKDLTSSNVVVKDPYKFLNKKERKDFLSSYNTIKDSIDTDINPANDVL 1383
DB 61 nfkvnlesddlipykdltsnvvvkdpykflnkkrdfklsyntykdsidtdinfandvl 120
QY 1384 GYKILSEKYSKSDLSIKKYYINDKOGENEKYLPLNNIETLYKTVNDKIDLFVTHLEAKV 1443
DB 121 gyykilseskysdlsikkyindkogenekylplnnietlykcvndkidlftvthleakv 180
QY 1444 LNTYKESNVEVKIKELNYLKTIDOKLADFRKNNNFVGIADLSTDYHNNLLTFELSTGM 1503
DB 181 lntykesnvevkiikelnylktldokladfrknnnfvgiadlstdyhnnlltfelstgm 240
QY 1504 VFENLAKTVLSNLDGMLQGLMNTISOHCYVKKCPONSQCFRHLDERECCCLNTYKOE 1563
DB 241 vfenlaktvlsnldgmlnsgmlnsqhcvcvkcpqnsqcftrhldereccclntykoe 300
QY 1564 DKCVENPPTCENNNGCGDADAKCTEEDSGSNGKRITCECTKPSYPLFDGIFCS 1618
DB 301 dkcvenpntcennngcgdadakcteedsngsklitcectkpsypfldgifcs 355

RESULT 6
ID AAY05833 standard; Protein: 361 AA.

XX AAY05833;

DT 02-AUG-1999 (first entry)

XX Merozoite surface protein MSP-1-42.

XX MSP-1; merozoite surface protein; malaria; vaccine;
KM protein engineering; protein expression; codon usage;
transgenic animal.

XX Plasmodium falciparum.

XX WO9920766-A2.

XX 29-APR-1999.

XX 20-OCT-1998; 98WO-US22225.

XX 15-MAY-1998; 98US-0085649.

XX 20-OCT-1997; 97US-0062592.

XX (GEN2) GENZYME TRANSGENICS CORP.

XX Chen LH, Meade H;

XX WPI; 1999-302742/25.

XX N-PsDB; AAX25587.

XX New modified recombinant nucleic acid sequences useful for producing
 PT malarial DNA vaccine
 XX
 PS Disclosure; Fig 2; 43pp; English.
 XX
 CC This present sequence comprises a 42 kDa C-terminal portion of
 CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
 CC important target for the development of a vaccine against
 CC Plasmodium falciparum. The C-terminal end of the sequence is
 CC modified to include a 6xHis tag. Nucleic acids (see AAX25386 and
 CC AAX25393) encoding MSP-1-42 have been modified according to a method
 CC of the invention in order to improve expression in mammalian cells
 CC and in transgenic animals. The invention provides modified
 CC recombinant nucleic acid sequences and methods for increasing the
 CC mRNA levels and protein expression of proteins that are difficult
 CC to express in cell culture systems, especially mammalian cell
 CC culture systems or in transgenic animals. The preferred difficult
 CC protein candidates for expression are those derived from lower
 CC organisms such as parasites, bacteria and viruses that have DNA
 CC coding sequences of high AT content or which have mRNA instability
 CC motifs or rare codons relative to the recombinant expression system
 CC to be used. The invention allows expression of MSP-1 in the milk
 CC of transgenic animals, and also provides a DNA vaccine comprising a
 CC vector containing the altered MSP-1-42 nucleic acid.
 XX
 SO Sequence 361 AA;

Query Match 22.3%; Score 1881; DB 20; Length 361;
 Best Local Similarity 100.0%; Pred. No. 1.9e-75;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1264 AVTPSVIDNLSKTENEEVLYLKPLAGVYRSLSKKOLENNVMTFNVVKDILNSRFNKR 1323
 DB 1 avtpsvidnlskteniyevelyikplagvyrslkqlennvmtfnvkvdkllnsrfnkre 60
 QY 1324 NFKNVLESDDLIPYKDLTSSNVVVKDPYKFLNKKERKDFLSYNYIKDSIDTDINPANDVL 1383
 DB 61 nfkvnlesddlipykdltssnyvvvdkpykflnkkekrdfklsyynyikdsiddinfandvl 120
 QY 1384 GYKILSEKYSKSDLSIKKTYNDKOGENEKYLPLNLNITETLYKTVNDKIDLFVTHLEAKV 1443
 DB 121 gykilsesksdlsikkyndkogenekylplnlmetlyktvndkidlfvthleakv 180
 QY 1444 LNTYKESNVEVKIKELNLYKTIDOKLADFEKKNNFVGIALSDYNNNNLTLFSLTGM 1503
 DB 181 lntyeksnvevkiikelnylktidokladfekknnfvgiadlstdynnnnltkflstgcm 240
 QY 1504 VFENLAKTVLSNLDGNLQGLNLSIQHCYVKKQCPONGSCFRHLDERECKCLNLYKQEG 1563
 DB 241 vfenlaktvlsnldgnlqglnlsgqhcvykkqcpngscfrhldececkclnlykqeg 300
 QY 1564 DKCVENRPTCNENNGGCDADAKCTEEDSGSNGKRTCECTKPDSPYLPFGCIFCS 1618
 DB 301 dkcvenrptcnennnggcdadakcteedsngskrtcectkpsdpylpfgcifcs 355

RESULT 7
 ID AAY09374 standard; Protein: 376 AA.
 XX
 AC AAY09374;
 XX
 DT 31-AUG-1999 (first entry)
 XX
 DE Modified merozoite surface protein MSP-1-42.
 XX
 KW MSP-1: merozoite surface protein; malaria; vaccine;
 KW protein engineering; protein expression; codon usage;
 XX transgenic animal; mutant.
 XX
 OS Plasmodium falciparum.

OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..15
 FT /note= "beta-casein signal peptide"
 FT Peptide 371..376
 FT /note= "6xHis tag"
 FT Misc-difference 197
 FT /note= "Asn in native MSP-1-42 (N181Q mutation)"
 FT Misc-difference 278
 FT /note= "Asn in native MSP-1-42 (N262Q mutation)"
 PN WO9920774-A2.
 PD 29-APR-1999.
 XX
 XX 20-OCT-1998; 98WO-0522226.
 XX
 XX 15-MAY-1998; 98US-0085649.
 PR 20-OCT-1997; 97US-0062592.
 XX
 PA (GENZ) GENZYME TRANSGENICS CORP.
 XX
 XX Chen LH, Meade H;
 XX PI
 XX WPI: 1999-288313/24.
 DR P-PSDB; AAX56008.
 XX
 PT Modified malarial protein for use in anti-malarial vaccines
 PS Example; Fig 11; 35pp; English.

CC The present sequence represents a 42 kDa C-terminal portion of
 CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
 CC important target for the development of a vaccine against
 CC Plasmodium falciparum. The sequence has been modified to include
 CC an N-terminal beta-casein signal peptide and an N-terminal 6xHis
 CC tag. In addition, N181Q and N262Q mutations have been introduced to
 CC eliminate 2 N-glycosylation sites. These modifications allow the
 CC MSP-1-42 protein to be expressed in the milk of transgenic mice.
 XX
 SO Sequence 376 AA;

Query Match 22.2%; Score 1869; DB 20; Length 376;
 Best Local Similarity 99.44%; Pred. No. 6.5e-75;
 Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1264 AVTPSVIDNLSKTENEEVLYLKPLAGVYRSLSKKOLENNVMTFNVVKDILNSRFNKR 1323
 DB 16 avtpsvidnlskteniyevelyikplagvyrslkqlennvmtfnvkvdkllnsrfnkre 75
 QY 1324 NFKNVLESDDLIPYKDLTSSNVVVKDPYKFLNKKERKDFLSYNYIKDSIDTDINPANDVL 1383
 DB 76 nfkvnlesddlipykdltssnyvvvdkpykflnkkekrdfklsyynyikdsiddinfandvl 135
 QY 1384 GYKILSEKYSKSDLSIKKTYNDKOGENEKYLPLNLNITETLYKTVNDKIDLFVTHLEAKV 1443
 DB 136 gykilsesksdlsikkyndkogenekylplnlmetlyktvndkidlfvthleakv 195
 QY 1444 LNTYKESNVEVKIKELNLYKTIDOKLADFEKKNNFVGIALSDYNNNNLTLFSLTGM 1503
 DB 196 lntyeksnvevkiikelnylktidokladfekknnfvgiadlstdynnnnltkflstgcm 255
 QY 1504 VFENLAKTVLSNLDGNLQGLNLSIQHCYVKKQCPONGSCFRHLDERECKCLNLYKQEG 1563
 DB 256 vfenlaktvlsnldgnlqglnlsgqhcvykkqcpngscfrhldececkclnlykqeg 315
 QY 1564 DKCVENRPTCNENNGGCDADAKCTEEDSGSNGKRTCECTKPDSPYLPFGCIFCS 1618
 DB 316 dkcvenrptcnennnggcdadakcteedsngskrtcectkpsdpylpfgcifcs 370

Query Match	Best Local Similarity	Score	DB	Length
Matches 353; Conservative	99.4%	22.2%	1869	376
	0; Mismatches	2; Indels	0; Gaps	0

Qy	1264	AVTSVIDNTILSKTENXEYLVLYAKPLAGYVRSJLKKOLENNWFVFNWVKDLNSFRKRE	1323
Db	16	avcpvsvidnllsktlenyevlylkplagvyrslkqjlenmntfnvkvdlinsfrkre	75
Qy	1324	NFKWVLESDDLPRVKDLTRSSWVVKDPRKFLNKKRDPFLSSVNYIKDSIDTDINFANVL	1383
Db	76	nfkvwlesddlprvkdltrsswvvykdkpflnkkrtokflssvnylksdtdinfandvl	135
Qy	1384	GYYKILSEKYSDDLSTIKRYINDKQGENEYLPFLNNIETLTKYVNDKIDLFLVHLKAV	1443
Db	136	gyyilsekyksdldsikkyindkgeneekyrlpflnnietlyktyvndkidlflvhlkav	195
Qy	1444	LNITYEESNVEVKIKELNYTKTQDKLADPKNNKNNFVGADLSDTYNNHNLTKFLSGM	1503
Db	196	lqyeyesnvevkielnylktlqdkladfkknntfvgiadlstidynhnnltkflsgm	255
Qy	1504	VFEKLAATVLSNLLDGLMGLMISOHOCVKKOCPONSGCFPHLDERECCCLINYKQEG	1563
Db	256	vfehlaktvlsnllldglmgmlisqhgcvkkypqpsngcfphldereccclinykqeg	315
Qy	1564	DKCEVNPNCNENNGCCDADAKCTEEDSGSNGKKTTCCTKPDSPVPLFDGIFCS	1618
Db	316	dkcevpnpncnenngccdadacteedsgsngkkttcctkpdspvplfdgifcs	370
RESULT	9		
AAAB37610			
ID	AAAB37610	standard; Protein; 262 AA.	
AC	AAAB37610;		
XX			
DT	27-FEB-2001	(first entry)	
XX			
DE	Merozoite surface protein-133.		
XX			
KM	Merozoite surface protein; protozoacide; vaccine; malaria.		
XX			
OS	Plasmodium falciparum.		
XX			
PN	W0200063245-A2.		
XX			
PD	26-OCT-2000.		
PF	20-APR-2000; 2000WO-GB01558.		
XX			
PR	20-APR-1999; 99GB-0009072.		
PR	13-MAY-1999; 99US-0311817.		
PR	25-MAY-1999; 99CA-2271451.		
XX			
PA	(MEDI-) MEDICAL RES COUNCIL.		
PI	Holder A, Birdsell B, Feeney J, Morgan W, Syed S, Uthaiplubul C;		
DR	WPI: 2001-015762/02.		
DR	N-PSDB; AAC68978.		
XX			
PT	Novel variants of the C-terminal fragment of Plasmodium merozoite		
PT	surface protein-1, useful as vaccines for treating or preventing		
PT	malaria		
XX			
XX	Example 5; Fig 15; 126pp; English.		
XX			
CC	The present invention relates to non-natural variants of a C-terminal		
CC	fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The		
CC	non-natural variants have reduced affinity for at least 1 antibody		
CC	capable of blocking a second antibody that inhibits the proteolytic		
CC	cleavage of Plasmodium MSP-1-4,2, and has the same affinity for at least		
CC	one third antibody that inhibits the proteolytic cleavage of Plasmodium		
CC	MSP-1-4,2, compared to natural MSP-1-1,9. The non-natural variants of the		
CC	present invention are useful for immunising a mammal against malaria, and		
CC	can be used to treat malaria. The present sequence is MSP-133 protein.		
XX			
XX	Sequence 262 AA;		
XX			

Query Match 15.8% Score 1333; DB 22; Length 262;
 Best Local Similarity 99.6%; Pred. No. 1.3e-51;
 Matches 261; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1264 AVFSPVIDNLSIKTENEXEVLTKPLAGVYRSRKOLENNVWFNVNKDILNSRFNRE 1323
 1 avtpsvdnlslsktenexevlytkplagvyrslkqlennvmtfnvnxkdlnsrfnre 60
 |||

QY 1324 NFKVNLSDILPYDNLSSNVVADPKFKLNKEKRDKFLSSYNTIKOSIDPDINADV 1383
 61 nfkvnlesdillykdlssnyvvdpkfllnkekrdkflssyntikdsidtdinfadv 120
 |||

QY 1384 GYXKLSKRYKSDSDSTKKYINDKOGENEKYLPLNNETLYKTWYNNKIDLFVHLEAKV 1443
 121 gyxklskryksdldstkkylndkgyenekylplnnelyktvndkldlfvhlleakv 180
 |||

QY 1444 LNTYKESNWEVKTREKLTNTKTIDKLADFKKNNNFVGIADLSTDYNNHNLTRFLSTGM 1503
 181 lntykesnwevktreklntktidkladfkknntfvgiadlstdyndhnltrflstgm 240
 |||

QY 1504 VFENLAKTVLSNLDGSLQGM 1525
 241 vfenlaktvlsnldgslqgm 262
 |||

RESULT 10
 AAP80547 ID AAP80547 standard; protein: 355 AA.
 XX
 AC AAP80547;
 XX
 DT 11-NOV-1990 (first entry)
 XX
 DE Polypeptide p190-3.
 XX
 KM Polypeptide p190-3; P. falciparum; merozoite; vaccine.
 XX
 OS Plasmodium falciparum.
 XX
 FH Key Location/Qualifiers
 FT Region 1..7
 FT /label-affinity peptide
 FT Region 144..355
 XX
 PN EP283829-A.
 XX
 PD 28-SEP-1988.
 XX
 PE 08-MAR-1988; 88EP-0103564.
 XX
 PR 19-MAR-1987; 87GB-0006599.
 XX
 PA (HOFF) F. HOFFMANN-LA ROCHE & CO.
 XX
 PI Ulrich C, Gentz H, Takacs B;
 XX
 DR WPI; 1988-272339/39.
 DR N-PSDB; AAN81151.
 XX
 PT New immunogenic polypeptides - derived from Plasmodium falciparum 190 kD
 PT surface antigen precursor, useful in vaccines, and encoding DNA sequences
 XX
 PS Claim 7, Page 28; 45pp; German.
 XX
 CC The polypeptide contains at least one epitope of the 190 kD precursor of
 CC the major merozoite surface antigen of P. falciparum. The polypeptide is
 CC immunogenic and can be used in vaccines to protect against malaria.
 CC Presence of the affinity peptide allows the polypeptide to be purified
 CC by affinity chromatography resins. Antibodies raised against the
 CC polypeptide are useful for diagnosis and for purification by affinity
 CC chromatography.

XX
 SQ Sequence 355 AA;

Query Match 10.8% Score 913; DB 9; Length 355;
 Best Local Similarity 80.6%; Pred. No. 5e-33;
 Matches 187; Conservative 9; Mismatches 14; Indels 22; Gaps 4;

QY 105 SGNRRTRNPSNNSSDSAKSYADLKHRYRNYLLTIKELKYQLDLNNHML-----TLCD 159
 106 tgnr---psvnnnd-----vnalesykrf-lpegdvtvsvessqtlcd 148
 |||

QY 160 NHHGFKYLIDYEETINELLYKINFFDLBRALNDVANCQIPFNKIRANLADV 219
 149 nhhgfkylidgyeetinellykintfydllralknnvcandycqipfnkiranldv 208
 |||

QY 220 LVFGYRPLDNIKONVGMEDYIKKNNKTENINELIBESKKTIDKNKATKEEKKLY 279
 209 lvfgyrpldnlkdnvgmedyiknnkktleninelieskkldknkatkeeeekky 268
 |||

QY 280 QAOYDLSTYNNKOLEAHNLSVLEKRRIDTLKKNENIKELLDKINIKNPPA 331
 269 qaydlstynkoleahnlsvlekrridtlkknenikellldkineiknpa 320
 |||

RESULT 11
 AAP82523 ID AAP82523 standard; protein: 219 AA.
 XX
 AC AAP82523;
 XX
 DT 12-DEC-1990 (first entry)
 XX
 DE 31-1 Repeated Delete.
 XX
 KM Malaria; stage-specific late schizont-merozoite antigen 31-1;
 KW vaccine; protozoan parasite; reptile.
 XX
 OS Plasmodium falciparum.
 XX
 PN EP254862-A.
 XX
 PD 03-FEB-1988.
 XX
 PE 20-JUN-1987; 87EP-0108867.
 XX
 PR 26-JUN-1986; 86US-0879076.
 XX
 PA (BEHW) BEHRINGWERKE AG.
 XX
 PI Shaw A, Humbert Y;
 XX
 DR WPI; 1988-030152/05.
 DR N-PSDB; AAN82176.
 XX
 PT Peptide(s) for prodn. of antiprotozoal vaccines - comprising
 PT surface antigen sequence with deleted reptile
 XX
 PS Disclosure; ; p: English.
 XX
 CC The sequence encodes a polypeptide which contains the unique
 CC sequence of the surface antigen of P. falciparum, but lacks the
 CC immunodominant repeat sequences which may allow the parasite to
 CC evade and decoy the immune system. The DNA for the stage-
 CC specific late schizont-merozoite antigen 31-1 in p31-1 (W08503725)
 CC was modified to delete the entire sequence coding for the reptile.
 CC The expression plasmid, 31-1 Repeated Delete, was introduced into
 CC E. coli K12 to produce the modified peptide. The transformant has
 CC been deposited in the Deutsche Sammlung von Mikroorganismen as
 CC E. coli K12 (958336c1) (p31-1 Repeat Delete). The peptide was used
 CC in compns. to treat and prevent protozoan parasitic infections.
 XX
 SQ Sequence 219 AA;

Db 189 daksyadlkhrvqnylfkikelpelldlenhnlclcdnhgfkylldgyeei 242

RESULT 14

AAW36103 ID AAP80546 standard; protein: 151 AA.

XX AAP80546;

DT 11-NOV-1990 (first entry)

DE Polypeptide p190-2b.

XX Polypeptide p190-2b; P.falciparum; merozoite; vaccine.

OS Plasmodium falciparum.

XX Key Location/Qualifiers

FT Region 1..5

FT Region /label=affinity peptide

PN EP283829-A.

PD 28-SEP-1988.

PF 08-MAR-1988; 88EP-0103564.

PR 19-MAR-1987; 87GB-0006599.

PA (HOFF) F. HOFFMANN-LA ROCHE & CO.

PI Ulrich C, Gentz H, Takacs B;

DR WPI: 1988-272339/39.

DR N-PSDB; AAN81150.

PT New immunogenic polypeptides - derived from plasmodium falciparum 190 kD

PS Claim 6; Page 27; 45pp; German.

CC the polypeptide contains at least one epitope of the 190 kD precursor of

CC the major merozoite surface antigen of P.falciparum. The polypeptide is

CC immunogenic and can be used in vaccines to protect against malaria.

CC Presence of the affinity peptide allows the polypeptide to be purified

CC by affinity chromatography resins. Antibodies raised against the

CC polypeptide are useful for diagnosis and for purification by affinity

CC chromatography.

CC

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX AAW36103;

AC 25-MAR-1998 (first entry)

DE P1MSPI(p19)A protein sequence.

XX Plasmodium vivax; merozoite surface protein; MSP1; p19;

XX Plasmodium falciparum; malaria; vaccine; immunity; epitope.

OS Plasmodium falciparum.

XX Key Location/Qualifiers

FT Region 1..95

FT Region /note="amino acids derived from P. falciparum MSP1 p19

FT Region /note="glycosylphosphatidylinositol anchoring sequence"

PN W09730158-A2.

PD 21-AUG-1997.

PF 14-FEB-1997; 97WO-FR00290.

PR 14-FEB-1996; 96FR-0001822.

PA (INSP) INST PASTEUR.

PI (UNY) UNIV NEW YORK STATE.

PI Barnwell JW, Longacre-Andre S, Mendis K, Nato F;

DR WPI: 1997-425033/39.

DR N-PSDB; AAT944550.

PT Recombinant protein containing the merozoite surface protein-1 p19

PT fragment - useful in anti-malarial vaccines, diagnosis and protein

PT purification

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Sequence 116 AA;

Query Match

Best Local Similarity 7.6%; Score 637; DB 18; Length 116;

Matches 113; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1526 NISOHCYKOCPPONSGFRHLDERECKCLNKKQESDKCVENPNPTGNNNGCQDADA 1585

DB 3 nlsqgcvkkkgcpensgcfrrhderECKCLNKKQEGDKCVENPNPTGNNNGCQDADA 62

QY 1586 KCTERDSSNKKRTTCECTKPDSPPLPFGTFCSSNPLFGTFLILMLILXSF 1639

DB 63 kcteedsgsngkktlcectkpdspyltfgltcssnflglsflllmlllysf 116

Search completed: October 27, 2001, 15:51:14
Job time: 9465 sec

RESULT 15
AAW36103 ID AAW36103 standard; protein: 116 AA.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2001, 15:51:48 : Search time 379.17 Seconds
(without alignments)
8180.589 Million cell updates/sec

Title: US-09-269-874-2

Perfect score: 4940

Sequence: 1 cgcagcgcgtatgaataatcat.....ttcatctatagatgatg 4940

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 730101 segs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_0601:*

1: /SIDSL1/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /SIDSL1/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /SIDSL1/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /SIDSL1/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /SIDSL1/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /SIDSL1/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /SIDSL1/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /SIDSL1/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /SIDSL1/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /SIDSL1/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /SIDSL1/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /SIDSL1/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /SIDSL1/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /SIDSL1/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /SIDSL1/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /SIDSL1/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /SIDSL1/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /SIDSL1/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /SIDSL1/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /SIDSL1/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT:*
22: /SIDSL1/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4938.4	100.0	4940	19	AAV21451
2	2837.6	57.4	4940	19	AAV5363
3	2706.2	54.8	5760	6	AAV50530
4	1262	25.5	5181	16	AAQ80911
5	1000	20.2	1950	18	AAV97956
6	989	20.0	1897	18	AAV97957
7	969.6	19.6	1896	18	AAV97957
8	744.2	15.1	1065	20	AAV56008
9	744.2	15.1	1065	20	AAV5586
10	742	15.0	1140	20	AAV56021
11	742	15.0	1140	20	AAV5593

12	719.8	14.6	1077	22	AAV68976	Merozoite surface
13	625.2	12.7	1088	20	AAV56009	Merozoite surface
14	625.2	12.7	1088	20	AAV5587	Merozoite surface
15	525.2	10.6	786	22	AAV68978	Merozoite surface
16	308.8	6.3	660	9	AAV81148	DNA sequence encod
17	308.8	6.3	1068	9	AAV81151	DNA sequence encod
18	246.6	5.0	354	18	AAV80403	PfMSP1(P19)A codin
19	246.6	5.0	354	18	AAV94550	PfMSP1(P19)A codin
20	243.4	4.9	737	6	AAV50355	Storage-specific,
21	240.2	4.9	668	9	AAV82176	31-1 Repeated dele
22	201.8	4.1	456	9	AAV81150	DNA sequence encod
23	201.8	4.1	462	9	AAV81149	DNA sequence encod
24	198.4	4.0	333	22	AAV68977	Merozoite surface
25	194.8	3.9	387	18	AAV80404	PfMSP1(P19)S codin
26	194.8	3.9	387	18	AAV94549	PfMSP1(P19)S codin
27	126.2	2.6	936	22	AAV58252	Oligonucleotide D1
28	126.2	2.6	936	22	AAV58254	Oligonucleotide D1
29	126.2	2.6	936	22	AAV58257	Oligonucleotide D1
30	126.2	2.6	936	22	AAV58259	Oligonucleotide D2
31	126.2	2.6	936	22	AAV58262	Oligonucleotide D2
32	126.2	2.6	938	22	AAV58255	Oligonucleotide D1
33	122.6	2.5	936	22	AAV58252	Oligonucleotide D1
34	122.6	2.5	936	22	AAV58254	Oligonucleotide D1
35	122.6	2.5	936	22	AAV58257	Oligonucleotide D1
36	122.6	2.5	936	22	AAV58259	Oligonucleotide D2
37	122.6	2.5	936	22	AAV58262	Oligonucleotide D2
38	122.6	2.5	938	22	AAV58255	Oligonucleotide D1
39	119.2	2.4	306	6	AAV50354	Storage-specific,
40	105.6	2.1	3399	17	AAV05868	Chicken leucocytos
41	99.4	2.0	165	18	AAV93729	DNA encoding signa
42	81.6	1.7	1685	16	AAV87587	DNA encoding leuco
43	70.6	1.4	3579	21	AAV70099	Plasmodium falci
44	65.2	1.3	8045	21	AAV55171	Neurospora crassa
45	60.6	1.2	5163	19	AAV20700	Cryptosporidium pa

ALIGNMENTS

RESULT 1	
AAV21451	standard; DNA: 4940 BP.
ID	AAV21451
XX	
AC	AAV21451;
DT	23-SEP-1998 (first entry)
DE	P. falciptarum modified gp190 DNA.
XX	
KW	gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
KW	monoclonal antibody; passive immunisation; parasite; ss.
XX	
OS	Plasmodium falciptarum.
OS	Synthetic.
XX	
FT	Key
FT	location/qualifiers
FT	10..4929
FT	/*tag= a
FT	/product= gp190
XX	
PN	W09814583-A2.
XX	
PD	09-APR-1998.
XX	
PF	02-OCT-1997; 97W0-EP05441.
XX	
PR	02-OCT-1996; 96DE-4040817.
XX	
PA	(BUA/) BUARD H.
XX	
PI	Bujard H, Pan W, Tolle R;
XX	
DR	WPI, 1998-240088/21.

DR P-PSDB: AAW54145.

XX Recombinant production of complete gp190/MSP-1 Plasmodium surface
PT protein - useful in anti-malaria vaccines, also stabilising genes by
PI reducing their AT content

PS Example 1; Fig 3c: 48bp; German.

XX This sequence encodes a modified Plasmodium falciparum gp190/MSP-1
CC (merozoite surface protein) which has a reduced AT content resulting in
CC a higher stability of the protein. Such a protein is useful in
CC vaccines against malaria or for producing monoclonal antibodies (for
CC passive immunisation). The complete gp190 protein can now be produced
CC outside the parasite and has, at least over extended regions, the native
CC pattern of folding. Larger amounts of the protein can be produced
CC recombinantly than would be possible using the parasites as source.

XX Sequence 4940 BP; 1669 A; 1156 C; 1060 G; 1055 T; 0 other;

Query Match 100.0%; Score 4938.4; DB 19; Length 4940;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 4939; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 cgcacgcgataaaaaatcattcttcctcgtgtcaattctgtttttatcataact 60
DB 1 cgcacgcgataaaaaatcattcttcctcgtgtcaattctgtttttatcataact 60
OY 61 cagtcgctgacccagcaaatccatcagaagcgtgttaagaactggaagcttgaagat 120
DB 61 cagtcgctgacccagcaaatccatcagaagcgtgttaagaactggaagcttgaagat 120
OY 121 gccgcctctaccggtatcacagcctgttccagaagagaagatgttctgaatgaaggagc 180
DB 121 gccgcctctaccggtatcacagcctgttccagaagagaagatgttctgaatgaaggagc 180
OY 181 agtggcagcgcgtgtacacacacacccgggttctaaagggtctgtgtagcggtggc 240
DB 181 agtggcagcgcgtgtacacacacacccgggttctaaagggtctgtgtagcggtggc 240
OY 241 tccggtgtgtctgtgctctgtgggttccgtcgcctccggcgagcggtgcatcagat 300
DB 241 tccggtgtgtctgtgctctgtgggttccgtcgcctccggcgagcggtgcatcagat 300
OY 301 ggcctcagtcgacgcggttcgcggagacagtcgaagaacacatccatctcaaacctc 360
DB 301 ggcctcagtcgacgcggttcgcggagacagtcgaagaacacatccatctcaaacctc 360
OY 361 agcagcttcgacgcgaagtcctacgcgcagcctcaagcaccgggtgagaatactctc 420
DB 361 agcagcttcgacgcgaagtcctacgcgcagcctcaagcaccgggtgagaatactctc 420
OY 421 actatcaaggagctgaagtaccacagctgttgcagcctcaatcatatgtgcacgt 480
DB 421 actatcaaggagctgaagtaccacagctgttgcagcctcaatcatatgtgcacgt 480
OY 481 tgtgataacattcagcttcacaaatctgattgacggtttaagagagatccaatgaac 540
DB 481 tgtgataacattcagcttcacaaatctgattgacggtttaagagagatccaatgaac 540
OY 541 ctgtacaagtgaatttctacttctgactgttaaggccaaactgaatgaagcttgcgc 600
DB 541 ctgtacaagtgaatttctacttctgactgttaaggccaaactgaatgaagcttgcgc 600
OY 601 aatgactatgtcaaatccatctcaatttgaagatcagagccaagattgagctatgt 660
DB 601 aatgactatgtcaaatccatctcaatttgaagatcagagccaagattgagctatgt 660
OY 661 aagaagttgtcttcgcatcgaagcctcgcgaacatcaagagaatgttggaag 720
DB 661 aagaagttgtcttcgcatcgaagcctcgcgaacatcaagagaatgttggaag 720
OY 721 atggaagattatataaaagaataagaacacatcgaacattgaagctgtatcga 780

DB 721 atggaagattatataaaagaataagaacacatcgaacattgaagctgtatcga 780
OY 781 gaatccaaaagaacattgacaaaataagaatgacaccaaaggagaagaagaagag 840
DB 781 gaatccaaaagaacattgacaaaataagaatgacaccaaaggagaagaagaagag 840
OY 841 tgtacacagccagtcagaccctgtccatctaacaacaagcttgaagaagccatac 900
DB 841 tgtacacagccagtcagaccctgtccatctaacaacaagcttgaagaagccatac 900
OY 901 ctcatcagcgtactgtgagaagcgcataagaccctcaagaagaatgaaatatcaaga 960
DB 901 ctcatcagcgtactgtgagaagcgcataagaccctcaagaagaatgaaatatcaaga 960
OY 961 ctgctcgcgaagaattgaattgaatctctccgcagcgaacattctggaacacct 1020
DB 961 ctgctcgcgaagaattgaattgaatctctccgcagcgaacattctggaacacct 1020
OY 1021 aacacgctgtgacagaagaacagaagatagagagcagagaagaatcaagaagatc 1080
DB 1021 aacacgctgtgacagaagaacagaagatagagagcagagaagaatcaagaagatc 1080
OY 1081 gccaaaaccattgaattcaacatagattctcttactgatacccttgaagctgagatc 1140
DB 1081 gccaaaaccattgaattcaacatagattctcttactgatacccttgaagctgagatc 1140
OY 1141 tacttgagagagaagaataagaatataagacatctccgcgaagtcgagacaagaatca 1200
DB 1141 tacttgagagagaagaataagaatataagacatctccgcgaagtcgagacaagaatca 1200
OY 1201 accgaaccattgaatatacccatgtgtgacgtaccctctgtctataacagatcac 1260
DB 1201 accgaaccattgaatatacccatgtgtgacgtaccctctgtctataacagatcac 1260
OY 1261 aacgcttcaacgagctcaatgaatgtcttcggtgaattcaaccctcgatatcga 1320
DB 1261 aacgcttcaacgagctcaatgaatgtcttcggtgaattcaaccctcgatatcga 1320
OY 1321 gaacctctgaagaatatactacacagacaatgagagaagaagttatacagaatcaag 1380
DB 1321 gaacctctgaagaatatactacacagacaatgagagaagaagttatacagaatcaag 1380
OY 1381 gagaagatcaaaattgagaagaagaataatgagatggaagaagaagttacgaagaacgc 1440
DB 1381 gagaagatcaaaattgagaagaagaataatgagatggaagaagaagttacgaagaacgc 1440
OY 1441 agcaaaagtctaaacgatatcactaaagatagaagaagctgtgaacgagatctatgat 1500
DB 1441 agcaaaagtctaaacgatatcactaaagatagaagaagctgtgaacgagatctatgat 1500
OY 1501 tccaatttcaacaataaataatgcagcctgcgaacttcgagaataatgattggaagaacgtgac 1560
DB 1501 tccaatttcaacaataaataatgcagcctgcgaacttcgagaataatgattggaagaacgtgac 1560
OY 1561 tcttacaagtcgagaagaacttgaacacacataatccttgcataccctgaagaattcgaag 1620
DB 1561 tcttacaagtcgagaagaacttgaacacacataatccttgcataccctgaagaattcgaag 1620
OY 1621 cataactcttgaagaagcttcaacaagaagcttgaatataagggagatctctcgcgagac 1680
DB 1621 cataactcttgaagaagcttcaacaagaagcttgaatataagggagatctctcgcgagac 1680
OY 1681 attgtgtgtgagaagaagcattgaatatacagaatccataagtaagatcgaaaacgag 1740
DB 1681 attgtgtgtgagaagaagcattgaatatacagaatccataagtaagatcgaaaacgag 1740
OY 1741 atcgagacgctgtgtgagaagaatcaagaagtcgaagaacgctgtgttgagaagaagatt 1800
DB 1741 atcgagacgctgtgtgagaagaatcaagaagtcgaagaacgctgtgttgagaagaagatt 1800
OY 1801 acaaaagacgaanaataaacccagatgagaagatccctggaggtctcgatatgttaaac 1860
DB 1801 acaaaagacgaanaataaacccagatgagaagatccctggaggtctcgatatgttaaac 1860

Db 1801 acaaaagcgaataaacaacagatgagaaagatcctgagctccgatatgttgaagtc 1860
Qy 1861 caagtgcaagagtgctctctcaatgaacaagattgatactccaagaagactcaatctat 1920
Db 1861 caagtgcaagagtgctctctcaatgaacaagattgatactccaagaagactcaatctat 1920
Qy 1921 ctgaaagacgtggaatttaaaacataataataatgtgcccgaatagtataaagcagagat 1980
Db 1921 ctgaaagacgtggaatttaaaacataataataatgtgcccgaatagtataaagcagagat 1980
Qy 1961 aagcaggaacatactactccatctgactcaagaagaagatagacaaactgaagtgttc 2040
Db 1961 aagcaggaacatactactccatctgactcaagaagaagatagacaaactgaagtgttc 2040
Qy 2041 atgcccgaagctcgagagctgatacaacgaagaagaagaacattaaactgaaagacag 2100
Db 2041 atgcccgaagctcgagagctgatacaacgaagaagaagaacattaaactgaaagacag 2100
Qy 2101 tcagataactccgagcctccacagaagaagataaacccgagctacccacagccc 2160
Db 2101 tcagataactccgagcctccacagaagaagataaacccgagctacccacagccc 2160
Qy 2161 ggaacaacagccggtltcagctctcgaagcgataagcgtgcagactcaagcacaagacag 2220
Db 2161 ggaacaacagccggtltcagctctcgaagcgataagcgtgcagactcaagcacaagacag 2220
Qy 2221 aagcagagacagcctccagctgcccagtgcccgttccagaagcttaagcccaagtgctaca 2280
Db 2221 aagcagagacagcctccagctgcccagtgcccgttccagaagcttaagcccaagtgctaca 2280
Qy 2281 ccaccagctcctgtgaataaacaagccgagaatgtgcagcaactggaactgaacttgaag 2340
Db 2281 ccaccagctcctgtgaataaacaagccgagaatgtgcagcaactggaactgaacttgaag 2340
Qy 2341 ctctatagcttccatgaatacatctcaactgtgcacaataatatactctgtctcacaagc 2400
Db 2341 ctctatagcttccatgaatacatctcaactgtgcacaataatatactctgtctcacaagc 2400
Qy 2401 actatgaacggaagaatctttaaacagtaacagataacccaaggaagagagataaactg 2460
Db 2401 actatgaacggaagaatctttaaacagtaacagataacccaaggaagagagataaactg 2460
Qy 2461 tccctctgtgataccactlgaactcgtgtltcaatataccagaacaactcccgltatgtat 2520
Db 2461 tccctctgtgataccactlgaactcgtgtltcaatataccagaacaactcccgltatgtat 2520
Qy 2521 tctatgttctgaatgacctcaacaatctctcttccaactgttcaatgagatatagaag 2580
Db 2521 tctatgttctgaatgacctcaacaatctctcttccaactgttcaatgagatatagaag 2580
Qy 2581 gagatggtctgaacactgttataaactcaagaacgaacgaagattaaagaactcttgag 2640
Db 2581 gagatggtctgaacactgttataaactcaagaacgaacgaagattaaagaactcttgag 2640
Qy 2641 gaaagctaaagaggtctccaactcgtgttaaaactctctctccaagcttccatgcaacactg 2700
Db 2641 gaaagctaaagaggtctccaactcgtgttaaaactctctctccaagcttccatgcaacactg 2700
Qy 2701 tctctcacactcaagacaagcccgaaagtgaagcgtctaaagcagacacttcaactcgacc 2760
Db 2701 tctctcacactcaagacaagcccgaaagtgaagcgtctaaagcagacacttcaactcgacc 2760
Qy 2761 aaccttaataactgaactgtttgagaacatctcgtctctctggaagaagataagac 2820
Db 2761 aaccttaataactgaactgtttgagaacatctcgtctctctggaagaagataagac 2820
Qy 2821 atctacccaagaactatgtggaacgaatcgtccggaactctctacgaagaagataactgaa 2880
Db 2821 atctacccaagaactatgtggaacgaatcgtccggaactctctacgaagaagataactgaa 2880
Qy 2881 gacagcgacacattctataacgagagcttcaactaactcgtgaatacttaagccgagtat 2940
Db 2881 gacagcgacacattctataacgagagcttcaactaactcgtgaatacttaagccgagtat 2940

Qy 2941 atcaactctttaagatgaatcttaacgtlaagaagcgtggaagagagacataaagctg 3000
Db 2941 atcaactctttaagatgaatcttaacgtlaagaagcgtggaagagagacataaagctg 3000
Qy 3001 aagaagacactgcaactgaactgactgacactgttacaacaagctacaactgaactgagaga 3060
Db 3001 aagaagacactgcaactgaactgactgacactgttacaacaagctacaactgaactgagaga 3060
Qy 3061 ctcttcgaagaagaagacagctcggaagltataagatgcagatcaagaagltgactctg 3120
Db 3061 ctcttcgaagaagaagacagctcggaagltataagatgcagatcaagaagltgactctg 3120
Qy 3121 ctcaagagcagcgttggaagaagaactcaactcactgacacatctcgaagaagctactgag 3180
Db 3121 ctcaagagcagcgttggaagaagaactcaactcactgacacatctcgaagaagctactgag 3180
Qy 3181 aacttctcagltgtcttcaacaagaagaagaagcgcgagatccgcgagacagagacact 3240
Db 3181 aacttctcagltgtcttcaacaagaagaagaagcgcgagatccgcgagacagagacact 3240
Qy 3241 ctggaagaacccaagaattcttctcaaacactacaagaagcctcgltcaagtatatatagtc 3300
Db 3241 ctggaagaacccaagaattcttctcaaacactacaagaagcctcgltcaagtatatatagtc 3300
Qy 3301 gagtcttctctctgaagaactctctcgaagagagcagcagacacagagataactagcc 3360
Db 3301 gagtcttctctctgaagaactctctcgaagagagcagcagacacagagataactagcc 3360
Qy 3361 agcctcgaagaacttcaagaactcgtgtltaagctcgaaagcagaactgaagaacacttgac 3420
Db 3361 agcctcgaagaacttcaagaactcgtgtltaagctcgaaagcagaactgaagaacacttgac 3420
Qy 3421 ctggaagaagaagaagctcagctactctctctagcgaactgcacacactgtatcgcgagctc 3480
Db 3421 ctggaagaagaagaagctcagctactctctctagcgaactgcacacactgtatcgcgagctc 3480
Qy 3481 aaggaagtcataaagaacaagaactacacccgcaatagcccaagcgagaataataacagac 3540
Db 3481 aaggaagtcataaagaacaagaactacacccgcaatagcccaagcgagaataataacagac 3540
Qy 3541 gtgaataacgcactlgaatcttacaagaagltcctcgtccttgaagaacagatgtcgccact 3600
Db 3541 gtgaataacgcactlgaatcttacaagaagltcctcgtccttgaagaacagatgtcgccact 3600
Qy 3601 gtggtgtctgaatctcgtgtccgacacacttgagacagcttcaaacctaagaagcctgcatct 3660
Db 3601 gtggtgtctgaatctcgtgtccgacacacttgagacagcttcaaacctaagaagcctgcatct 3660
Qy 3661 actcatgtcgaagccgagctccaatacaattaccacatctcgaagaagctgcagatgagctc 3720
Db 3661 actcatgtcgaagccgagctccaatacaattaccacatctcgaagaagctgcagatgagctc 3720
Qy 3721 gatgacgtcatcatgtgtcctatctctcgcgagagcggagagactacatgacactcgcg 3780
Db 3721 gatgacgtcatcatgtgtcctatctctcgcgagagcggagagactacatgacactcgcg 3780
Qy 3781 caggtgttcaacccgtlgaagcgtgtcaactccttcgltgattgataaactctgtccaanaatc 3840
Db 3781 caggtgttcaacccgtlgaagcgtgtcaactccttcgltgattgataaactctgtccaanaatc 3840
Qy 3841 gagaaagcaatcgaagtgctctatctatctgaaacctcgtggaagcgttcataggtcttcag 3900
Db 3841 gagaaagcaatcgaagtgctctatctatctgaaacctcgtggaagcgttcataggtcttcag 3900
Qy 3901 aaacagctggaagaataacgtatgacacttcaatgtcaacgctgaaagacatcttgacagc 3960
Db 3901 aaacagctggaagaataacgtatgacacttcaatgtcaacgctgaaagacatcttgacagc 3960
Qy 3961 cgctttaaagaagaagaataattcaagaagcgtcttggaagcagactgtatctccataaa 4020
Db 3961 cgctttaaagaagaagaataattcaagaagcgtcttggaagcagactgtatctccataaa 4020

```

QY 4021 gacctgacctctctaactagctgttcaaggagccatacaagtctcctaataagaagaag 4080
    |||
DB 4021 gacctgacctctctaactagctgttcaaggagccatacaagtctcctaataagaagaag 4080
QY 4081 agggataaattctctagttacatactatacaagagctccatcgacacgatacaat 4140
    |||
DB 4081 agggataaattctctagttacatactatacaagagctccatcgacacgatacaat 4140
QY 4141 ttccgataatgtctgtctgggtatatacaagatcctgaagcggaaaaataaagtctgacctt 4200
    |||
DB 4141 ttccgataatgtctgtctgggtatatacaagatcctgaagcggaaaaataaagtctgacctt 4200
QY 4201 gactctataaataaataatacaacgataaagcgaagcggaatgtaaaaaatctgccttc 4260
    |||
DB 4201 gactctataaataaataatacaacgataaagcgaagcggaatgtaaaaaatctgccttc 4260
QY 4261 ctgataaatacgcgaacacctgtacagaagcgtgacgacaaatcgacctctgtaatt 4320
    |||
DB 4261 ctgataaatacgcgaacacctgtacagaagcgtgacgacaaatcgacctctgtaatt 4320
QY 4321 caccgtgagggccaaagctctcaactactactacgagaagagcaatgtggaatttaaatc 4380
    |||
DB 4321 caccgtgagggccaaagctctcaactactactacgagaagagcaatgtggaatttaaatc 4380
QY 4381 aaggagctgaactactactcaaaaacaaacccaagaagcgtgcagattccaagaataaac 4440
    |||
DB 4381 aaggagctgaactactactcaaaaacaaacccaagaagcgtgcagattccaagaataaac 4440
QY 4441 aattctgtcgcgaatttgcgaacctgtctacagattataacacacaaatctctgaccaa 4500
    |||
DB 4441 aattctgtcgcgaatttgcgaacctgtctacagattataacacacaaatctctgaccaa 4500
QY 4501 ttctgtctcactgtgcatgtgttccgaaaacctgcgccaagaagctgtgagcaattgtctc 4560
    |||
DB 4501 ttctgtctcactgtgcatgtgttccgaaaacctgcgccaagaagctgtgagcaattgtctc 4560
QY 4561 gacggcaaacctgcgacggagctgtgacacatctcccgacacaaatgtgtgaagaacagctgc 4620
    |||
DB 4561 gacggcaaacctgcgacggagctgtgacacatctcccgacacaaatgtgtgaagaacagctgc 4620
QY 4621 ccccaagaatacggctgttctcaagcatctgacgacgagcgaagatgtcaagtgtctctg 4680
    |||
DB 4621 ccccaagaatacggctgttctcaagcatctgacgacgagcgaagatgtcaagtgtctctg 4680
QY 4681 aactacaacaagaagagataatgtgtgtgagaacccaacacctctgcaatgaataaac 4740
    |||
DB 4681 aactacaacaagaagagataatgtgtgtgagaacccaacacctctgcaatgaataaac 4740
QY 4741 aatggtcggtgtgtgacgacgataatgtacacccgagaagacagcggtcttaacggaag 4800
    |||
DB 4741 aatggtcggtgtgtgacgacgataatgtacacccgagaagacagcggtcttaacggaag 4800
QY 4801 aaatcacaatgcgagtgatcaagaagccgacacctctcttgaagggatttttttgc 4860
    |||
DB 4801 aaatcacaatgcgagtgatcaagaagccgacacctctcttgaagggatttttttgc 4860
QY 4861 tccagcttaattctcctgagcatcctctcctgctgacatccatgatacctgtgacagc 4920
    |||
DB 4861 tccagcttaattctcctgagcatcctctcctgctgacatccatgatacctgtgacagc 4920
QY 4921 ttcatctaataatgatcatgagc 4940
    |||
DB 4921 ttcatctaataatgatcatgagc 4940

```

RESULT 2

AAV35363

AAV35363:

23-SEP-1998 (first entry)

XX

DE P. falciparum gp190 DNA.

KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
KW monoclonal antibody; passive immunisation; parasite; ss.

OS Plasmodium falciparum.

FH Key Location/Qualifiers

FT CDS 10..4929

FT /tag=a

FT /product= gp190

XX MO814583-A2.

XX PD 09-APR-1998.

XX PF 02-OCT-1997; 97WO-EP05441.

XX PR 02-OCT-1996; 96DE-4040817.

XX PA (BUA/) BUJARD H.

XX PI Bujard H, Pan W, Tolle R;

XX DR WPI; 1998-240088/21.

XX PS P-PSDB; AAW54145.

PT Recombinant production of complete gp190/MSP-1 Plasmodium surface
PT protein - useful in anti-malaria vaccines, also stabilising genes by
PT reducing their AT content

PS Example 1; Fig 3c; 48bp; German.

CC This sequence encodes the Plasmodium falciparum gp190/MSP-1 (merozoite
CC surface protein). This gene is used in a method for stabilising the
CC gene sequences by reducing the AT content. Such products are useful in
CC vaccines against malaria or for producing monoclonal antibodies (for
CC passive immunisation). The complete gp190 protein can now be produced
CC outside the parasite and has, at least over extended regions, the native
CC pattern of folding. Larger amounts of the protein can be produced
CC recombinantly than would be possible using the parasites as source.

XX Sequence 4940 BP; 2196 A; 597 C; 687 G; 1460 T; 0 other;

Query Match 57.4%; Score 2837.6; DB 19; Length 4940;

Best Local Similarity 73.4%; Pred. No. 0; Mismatches 1314; Indels 0; Gaps 0;

Matches 3626; Conservative 0; Mismatches 1314; Indels 0; Gaps 0;

QY 1 cgaacgcatgaaataattctctctctgctcatctctgttttatacataact 60

DB 1 cgaacgcatgaaataattctctctctgctcatctctgttttatacataact 60

QY 61 cagtcgtgacccagaaatcctatacagagctgtgttaagaacgtgaagcttggaaat 120

DB 61 caatgttaacacagaaagtatacaagaactgtcaaaaactgaagacttgaagaat 120

QY 121 gccgtcttaacggatlaagcctgttccagaagaggaagatgtgtcgaatgaagagcg 180

DB 121 gcagattgacaggtlaagttatattcaaaagaaaaatgtaataaagaaagaa 180

QY 181 agtgcacgacgcttatacaacacacacccggtttcctaagggtctgtcagcgtgtgc 240

DB 181 agtgcacacgcttatacaacacacacacacacacacacacacacacacacacacac 240

QY 241 tccagtggtctgtgtcctctggtgttcgttcgctccggtcggcagcggtgcacagat 300

DB 241 tccagtggtctgtgtcctctggtgttcgttcgctccggtcggcagcggtgcacagat 300

QY 301 ggtcagtgacagcggttccggaacagtcgaagaacacacacacacacacacacac 360

DB 301 ggtcagtggtctcagtggttcaggaataatcaagacgtaacaaatcttcagataatca 360

OY	361	agcgatccgaagcgcgaagctccctacgccgacctcaagcagcgatgagaactatccctc	420
Dp	361	agtgatccaatgcataactcttaacgtctgatttaaaacacagatgcgaattacttgta	420
OY	421	actcaaggagcgtgaagatcccaagctgtgttcgacctacacatcatgtcgacgtg	480
Dp	421	actcacaagaacctcaaatatccccaactcccttgatttaactcaatcatgttlaacttg	480
OY	481	tgtatacaatcatgtgcttccaatatccttgatgacggtttaaagaaatgatactaac	540
Dp	481	tgtatacaatcatcgtgttccaatatcttaattgatgagatagaagaataatgaata	540
OY	541	ctgtacaagttgaattctactctgcactgtgctaaggccaactgaaatgacgtgtgc	600
Dp	541	ttatataaatttaacttttatttgatttataagcgcaaaattaaatgattgatgtct	600
OY	601	aatgacattgtccaattccattccaatttgaagatcagcgccaagagtttgacgtatg	660
Dp	601	aatgattatctgccaataaacccttccaactcttaaaattcgtgcaaatgaaatgacgtact	660
OY	661	aagaagttgtctcgtgatttcgcaagcctctcgcacaatcaataaggaacatgtggaag	720
Dp	661	aaaaacttggttcggtatagaaaacattagacaaatattaaagatcaatgtagaaa	720
OY	721	atggaagattatataaaagaataaagaacaccatcgcgaacattaacgagctgtatga	780
Dp	721	atggaagattacattataaaaaataaaaaacatagaanaatataatgatttaattgaa	780
OY	781	gaatccaaaagaagccatagagcaaaaataagatctgcaccagaagaggaagaagaag	840
Dp	781	gaagaatagaanaaanaattgtataaataaagaatgacaactaaagaagaagaaaaaaa	840
OY	841	cttgacaagggcccgatgcagcgcctgcacataaacaacagctggaagcccatatc	900
Dp	841	ttataccagctcaatagtctctcttatttaacaataaacaattgaagaagcacataat	900
OY	901	ctcatcagcgtaactgtgagaagcgcatagacacccctcaagaagaatgaaatacaaa	960
Dp	901	ttataaagcgttttagaanaaacgattctgacactttaaanaaaaatgaaacattaga	960
OY	961	ctgtctgacaagaattgaattgaagtctctccgcagccaactctcgggaacccct	1020
Dp	961	ttactctgatatagaanaatgaaattlaaaaatccccccgcggccaattctggaattaca	1020
OY	1021	aacgcgctgtcgtgacaagaacaagaagctatagaagcagcagaagaagatcaaaaga	1080
Dp	1021	atactctccttgataaagaacaaaaaattcggggaacgcgaanaagaataaagaat	1080
OY	1081	gccaaaaacctaagtlcaacatagatctctctcttactgataccctctgagctgga	1140
Dp	1081	gccaaaacttaatttaatttatgttagtttaattactgataccactggaattgaata	1140
OY	1141	tacttgagagagaagaatagaatatagacatctccgccaagtctgagacaaaggaat	1200
Dp	1141	tatttaagaaaaaaaataaataatctgtataagtgcagaagtggttgcacaaggaat	1200
OY	1201	accgaacctaatgaatccccaatgtgtgagctaccctctcttataagatatcaac	1260
Dp	1201	actgaacccaatgaatctccaatgtgagttacttacccttctgcataaacaagatat	1260
OY	1261	aacgctctcaacgagctcaatagctcttcggttgacttgattaacccctcogatata	1320
Dp	1261	aatgtcttaaatgaacttaattcttctgtgtgttaatttaacatcattgattata	1320
OY	1321	gaaccccttaagatatctcaacagaacatltgagagaagaagttatacaagaaatca	1380
Dp	1321	gaaccaaagtaaaaacatatactactgtataagaagaaaaaattcaatgtaaatt	1380
OY	1381	gagagagaatcaaaattgagaagaagaatttggagtgagacaagaagaagtttca	1440
Dp	1381	gaaaaaatttaaatatgaaaaaaaaaattggaatctgtataaaaaattcttcaaga	1440
OY	1441	agcaaaagcttaaacgatatcactaaagaagttgaanaagctgtgaacgagatcat	1500

[illegible]

Dd	2521	tcatagtttgtatgacttaacaatagtttataccaactattatggaattatgaaaaa	2560
Qy	2561	gagatggtctgcaaacctgttataaactcaaaagacaagacaagatlaagaacctctgag	2640
Dd	2581	gaaatggtctgtaatcttataacttaacttaagatgaatagcacaataatttaatttagag	2640
Qy	2641	gaagctaaagaaggtctccacctgtgttaaaactctctcttccagctccatgcaaccactg	2700
Dd	2641	gaagcgaaaaaagatataccactctgttaaaaactcttccagttcatcatgcaaccattaa	2700
Qy	2701	tcctcccaactctcaagccagcccggaagtgaggtctaaagaaagacacccctccaccgccc	2760
Dd	2701	tcattacaacctctcgatataaaccggaagtgaagtcgcaaatgagatatacatcatcttcaa	2760
Qy	2761	aaccttaataactccactgynaactgttttgagaacatctgttcctccgagaaagataagac	2820
Dd	2761	aatttgataataagttttaaataattttatgaanaacaactctgagctctgtgaaaaaacaanaat	2820
Qy	2821	atctaccaagaacttatttgcagcaaaaatcgctccggaacctctctacagagaatactgaaa	2880
Dd	2821	atatacccaagatataaaggtccaaaaagtagtgaanaaacttttgaanaagatataaaa	2880
Qy	2881	gacagcgacacattcttaacgagagcttcacactaacttgtagaactcctaaagcgatgat	2940
Dd	2881	gatagtgtacaacttttbaatgatacctcttcaacaattctgttaaaactcaaaagctgagat	2940
Qy	2941	atcaactctctttaaacyatgaatctaaacytaagaagcttggaagagagacaataaagctg	3000
Dd	2941	attaatcatctgaaatgagtgtgaatccaanaaggaagaatlagagaagatataataaata	3000
Qy	3001	aagaaagacactgcaacctgagcttcgacctgtacacaaagttacaaactgaaactgagaga	3060
Dd	3001	aaaaaaactttacagttatacatcttgtatttataataataataataataaattagaaga	3060
Qy	3061	ctcttcgcgaagaagaagagacagctccgcaagtataagaatgcagataagaagttagactg	3120
Dd	3061	ttatttgtataaagaagaaacaggttgynaataataaaatgynaatlaaaaaacttactta	3120
Qy	3121	ctcaagagacgagcttgaaagcaaacactcaactcaactcaaacctcgaaaacagttactgag	3180
Dd	3121	ttaaaagacaatattagatacaaaaattgaaatctcaacttaataaccgaagcatgtattaca	3180
Qy	3181	aactctcagttgtctctcaacaagaagaagaagccgagatctgcggaagacagaaacact	3240
Dd	3181	aactctctgtttcttctttaaacaanaaaagaagctgtaaatlagcagaacctgynaacacaa	3240
Qy	3241	ctggagagaacacacgaatctctctcaaacactcaaaaagcctctgtaagtattataatgagc	3300
Dd	3241	ctagaanaaacaaaaataataattgaaacattataaaaagagcttgttaaatattataatggt	3300
Qy	3301	gagctctctctctgaaagactctctccggagagagatccagaaccgagataaactacgccc	3360
Dd	3301	gaatcatcatctcatataaaaacttttaagtgaagaatcaatccaacagaagatataatgccc	3360
Qy	3361	agcctcgagaactctcaagctcctgtctgaagctcogaagcagaactgaaagacaacctgac	3420
Dd	3361	agtttaagaaaactttaaagattaaatagaagaagaaattaaagataatattaat	3420
Qy	3421	ctggagagaagaagcgcagctctacccctcttagcgagatgataccactgatactgcggagtc	3480
Dd	3421	ctagaanaagaanaaataatcatataactcatcaagtggtattacatcatattatgtcgaaatla	3480
Qy	3481	aagaaagctcatlaagaacaagaactacacccggaatlagcccaagcgagaaatatacagac	3540
Dd	3481	aagaagaatataaaaaataataatattacaggttaattctccaaagtgaanaataatacggat	3540
Qy	3541	gtgaataacgactgynaactcttcaagaagctcttcgctcgtaagagacaagatgctgcgac	3600
Dd	3541	gttaacaatgcatragataacttcaaaaaaattctccgaagagacagatggttgcaaca	3600
Qy	3601	gtggtgtctgaaactgtgctccgacacactgagcagttccaactaagaagctgcgact	3660
Dd	3601	gtcttaagtgaaagttgtagtctcgacacacttaggaacaagctcaacaaanaaacacgactca	3660

QY	3661	accatctgcgaagccgggtccaatataatccaaatctccagaacgttcgaatgaggtc	3720
Db	3661	actcatgtaggagagagctctaaacaatacaacaatctcaaaaatgttcgatgtgaagta	3720
QY	3721	gatgcagcatctgtgcctatcttcctgcgagagcgagagactacgatgacctgcgc	3780
Db	3721	gatgcagcatctagtaactatatttggagaatccgaagaagatattagatgtttaga	3780
QY	3781	caagtgtgcacccggttgagcgtgctaactccttcctcgatattgataacatctcgtccaaatc	3840
Db	3781	caagtagtaatacagagagagacgaatgaactcccttcctgtaattgatacaatacttctaaat	3840
QY	3841	gagaacggaatacgaagagcctctcttgaaacccctgcgcgagcgctctatagcgctcgaag	3900
Db	3841	gaataatgatactgtggttcttatattatataaaaccttgcgcagtggttataagaagtlttaa	3900
QY	3901	aaacagcgtggaataaacygtatgaccttcaatgltcaacagtgtaagagacattctgaacgc	3960
Db	3901	aaacatattagaataaataacgcttatgacatttaagttaatgttaagatattttaattca	3960
QY	3961	cgctttataagagaganaatttccaagaacgtcttggagagcgacttgattcccatataa	4020
Db	3961	cgatttaataacgtygaanaattccaanaagttttagaatcagatattcaatccatataa	4020
QY	4021	gacctgaacctctctaaactcgtctgtccaaggaaccaataaagtctcccaataaagagag	4080
Db	4021	gatttaacatacgaagtaattatgtcttccaagaagatccatataaattctcttaataagaanaa	4080
QY	4081	aggagataaattctgcctgaatttcaactatataccaagagctccatcgcgaacggatacat	4140
Db	4081	aggagataaattcttaagcagttataatlatatlaagattcaatagatagcagatataat	4140
QY	4141	ttcgctaatgtagtgcgtggggtatctcaagatccctgagaganaataatacagctgcacct	4200
Db	4141	tttgcgaatgtagtgccttgatataataaataatactccgaanaataatacaatccagattta	4200
QY	4201	gactctatataaagatataccaacgataagcaagcgagaaatgtaaaaatctgcacctc	4260
Db	4201	gactcaatataaanaatatataccaacgaacaaagtygaanaatgtaaaaataactccctct	4260
QY	4261	ctgatatcaatccgaacaccgctgtacaagacgtgtaagaganaaataatgcacctcttgatatt	4320
Db	4261	tttaacaatatgtgacctctataataaacgcttaatagataaaaatgcatttatctgtcaatt	4320
QY	4321	caacctggaagccaaggtgcctcaactaacttaccagagaagagcaatgtggaagttaaaatc	4380
Db	4321	catatagaagcaaaaagttcttaaatattatatacatgtaagaaatcaaacgttagaagttaaata	4380
QY	4381	aagagacgtgaactccctcaanaacatccaagacagctgcgcgatttccaagaanaataac	4440
Db	4381	aagaagacttaacttcaaaaacaatccaagaanaaatgtgcagatttcaaaaaaataaac	4440
QY	4441	aattctgtcgaattgtgacagccgtctacogattataacacaanaaatctctgcacaag	4500
Db	4441	aattctgtggaattgtgtttatataccaagattataacataacttaacttatgtgaagaag	4500
QY	4501	ttctctgtccactgcagatggtgtctcgaaaacctgcgcacaaacaggtgcagcaatctgctc	4560
Db	4501	ttctctgtatacaggtatggttttttgaaaactctgtctaaaacctgtttatcttaatttctt	4560
QY	4561	gagcgcaacctgcagggcagctgcgtacaactcccgacaccaaattggtgtgaagaaacgtgc	4620
Db	4561	gatgtgaacctgtcgaaggtatglttaaacatttcaacaacaccaaattggtgtataaaaacaatgt	4620
QY	4621	ccccagataagcgagctgttttcagacatctcgtgacgacgagcgagagtygaagtygtctcgt	4680
Db	4621	ccccaaatctcgtatggttttcagacatttaagtgaagaagaaatgttaaatgtttattat	4680
QY	4681	aactatacaacaagaagagataaagtgtcgtggtggaacccaacacctactctgcaatggaac	4740
Db	4681	aattatacaacaagaagtygtataaattgtgttgtaaaatccaatctctactctgtgaagaaat	4740

Oy	4741	aatgctgggtgtgagccgcgaatgctaaatgacccgggaagacagcgctcttaacggaaag	4800
Oy	4741		4800
Db	4741	aatggttggaatgtagacgcagatgccaatgtaaccgaagaagtccagtgtagcaacggaaag	4800
Oy	4801	aaatcacatgtgaagtgtaactaagccgacactctatccactcttcgacggagatttttcg	4860
Oy	4801		4860
Db	4801	aaatcacatgtgaagtgtaactaagccgacactctatccactcttcgacggagatttttcg	4860
Oy	4861	tccaagctctaatcttcctgagcatctctctctctgcatccatcatgcatcgtacagc	4920
Oy	4861		4920
Db	4861	agctctcttaactcttcttgagaaatcatcatcttattatatactatcatgttaatatatacagt	4920
Oy	4921	ttcatcctaataagatcgatcg	4940
Oy	4921		4940
Db	4921	ttcatcctaataagatcgatcg	4940
RESULT 3			
AAN50530			
ID	AAN50530 standard; DNA; 5760 BP.		
XX	AC	AAN50530;	
XX	DT	30-SEP-1991 (first entry)	
XX	DE	Sequence encoding the P195 protein of Plasmodium falciparum.	
XX	KW	Malaria vaccine; epitope; antigen; immunogen; ss.	
XX	OS	Plasmodium falciparum.	
XX	FH	Location/Qualifiers	
XX	FT	216..5179	
XX	FT	CDS	/*tag= a
XX	PN	EP154454-A.	
XX	PD	11-SEP-1985.	
XX	PF	21-FEB-1985; 85EP-0301173.	
XX	PR	26-SEP-1984; 84GB-0024340.	
XX	PR	22-FEB-1984; 84GB-0004692.	
XX	PR	21-FEB-1985; 85GB-0004429.	
XX	PA	(WELL) WELLCOME FOUNDATION LTD.	
XX	PI	Holder A, Sandhu J, Odink K, Lockyer M, Riveros-Moreno V;	
XX	DR	WPI: 1985-224845/37.	
XX	DR	P-PSDB: AAP50777.	
XX	PT	Cloned DNA sequence encoding plasmodium falciparum protein -	
XX	PT	useful for expressing the protein for use in vaccines against	
XX	PT	malaria	
XX	PS	Claim 2; Fig 1; 51pp; English.	
XX	CC	The sequence encoding the P195 protein of Plasmodium falciparum	
XX	CC	(AAN50530) and a peptide comprising at least one of its epitopes	
XX	CC	(see AAP50777) are claimed. Also claimed is a vaccine for inducing	
XX	CC	immunity to malaria comprising the novel peptide or P195 or a	
XX	CC	peptide comprising at least one epitope when derived from the new	
XX	CC	DNA sequence, together with a carrier.	
SO	Sequence 5760 BP; 2565 A; 630 C; 725 G; 1840 T; 0 other;		
Query Match 54.8%; Score 2706.2; DB 6; Length 5760;			
Best Local Similarity 72.1%; Pred. No. 0;			
Matches 3587; Conservative 0; Mismatches 1343; Indels 45; Gaps 3;			
Oy	10 atgaatacatcttctctctgtcatcttctgttttataatcaataactagtcgty 69		

Db	216	atgaagatcatattcttttatgttcaattctctttttattatataacaaatggtga	275
Oy	70	accacgatcccttcagggcgtgtttaaanaacttgaagctcttgagaatgcccctt	129
Db	276	acacatgaagttctcaagaacttgcataaaaaactgaagcctttgaagaatgcattgt	335
Oy	130	accgatatacagctgttccagaagagaaatgtgtctgaatgaaggagatgagatgagcag	189
Db	336	acagttatagttattttcaaaaggaaaaatgttatlaaatgaaggagaaacagtgaaaca	395
Oy	130	gaccttacaacagcacaccgggtcttaaaaggtctgtgtacggctgtgctcogtvgg	249
Db	336	gctttcaacactagtaaacctgttcaaaaggtctcagttgcagttgagttcaggtgagc	455
Oy	250	tcgtgtgctcttgggggtttccgtcgcctccggggagaggtggcactcaggttgcaggt	309
Db	456	tcaagttgtcttaaggtggtccagttgtctcagggtggtccagttgtctcaggtggtccagtt	515
Oy	310	gcaagcgcggtttccgggaacagtcgaagaccacacacacacacacacacacacacacac	369
Db	516	gtctcaggtgtgttcaggtaatttcaagaacgtaacacatcccttcagataaatccaagtattca	575
Oy	370	gaagccaagctctacgcgcgaactcaagcacagagtgagaactatccctccatacga	429
Db	576	gatgtaaatcttcagctgtgtttaaaacacagagtcgaataattctgttactatcaaa	635
Oy	430	gagctgaagtaaccacagctgtgtccagctccacatcatatgtctgaactgtgagactgtgatac	489
Db	636	gaacctcaaatctcccaactcttgtatttbaactacatcatatgttcaacttgtgtgataat	695
Oy	490	attcatgcttcaaatatctcgatttgcggttgaagagatcaatgaactccctgtgacaag	549
Db	696	attcatgtttcaaatatcttgaattgtatgtatagaagaataatgaattatataataa	755
Oy	550	ttgaattctactcgaactgtgtctaagggccaactgaatgcagtttgcgcgaatgactat	609
Db	756	ttaaacttttattgtattatttaagagcaaaattaaatgactgtgtgcatagtattat	815
Oy	610	tgtaaatctcatctcaatttgaagaalcatgagagccaagatgtgacgttatgaagaattgt	669
Db	816	tgtaaaatcccttcaactcttaaatcttgcgcgaatgaaatagagacttacttataaaaaact	875
Oy	670	gtcttcgagatctgcgaagcctctccgacaacatcaagagacaatgtgtggagaatgagaat	729
Db	876	gtgttcgagatctagaaaaaccattagacaatatataagaataatgtcggaaaaatggaagat	935
Oy	730	tatatlaaanaagaaatgaagaccatctgaataatcaatcaagcgtgtatctgaataatcccaa	789
Db	936	tacattaaaaaaaataaaaaaacattagaaaatataatgaattatgaagaagaatgaag	995
Oy	790	aagaccatagacaanaataaagaatcaacccaagaggaagaagaagaagaatgtgtacag	849
Db	996	aaaacaattgtttaaaaaatagaatgtcaacttaagaagaagaagaagaagaagaatlaaacaa	1055
Oy	850	gcccagtaagcaccgttcacatctatacaaacagcttgaagaagcccatcaactcaatccgc	909
Db	1056	gtccaatgtctcttctatttcaataaacaattatgaagaagcccatcaatttaataagc	1115
Oy	910	gtactgagaagcgcataagacacccctcaagaaatgaataatcaataaagaactgtgtcac	969
Db	1116	gttttgaanaaagtattgtgcactttlaaaaaaaaatgaaaacatttaaggaattactgtat	1175
Oy	970	aagattaatgaattaaatgaatctctccgcagcgaactctcgggaaacacccctaacacgctg	1029
Db	1176	aagatlaaatgtgaattlaaaaaatcccccacgcgcgaatcttgcggaataatcacacaatattct	1235
Oy	1030	ctggaacaagaacaagaagatagagagagcagcgaaagagatataaagagatcgcgcaaac	1089
Db	1236	cttgataagaacaacaaaaacgagagacaagaaaaagaataaagaagaattgtgccaacct	1295
Oy	1090	attaagtccaacatagatctctcttactgatacccttggagctgtgaatctattgtga	1149

D	1296	atuaatttaataatgtaagtttaatttaactgataccaattggaattgaatactatttaaga	1353
Q	1150	gagaagaataagaatalagacatcccgccnaagtcgagacaaaggaatlcaaccgaact	1209
D	1356	gaaaaataataaattatggtataagtcgaaggttgaaacaaaggaatcaactgaacc	1415
Q	1210	aatgaatatacccaatggtgtagagtaacctgtcttataaagataatacaacagcttc	1269
D	1416	aatgaatatacccaatgtagtactatccctgttcaataacgaatatacaaatgtctta	1475
Q	1270	aacgagctcaatagcttcggttgactgattaaaccccttgattatagcaagaacccct	1329
D	1476	aatgaacttaattcttttgtgtgttaataatccaattgattatacaaaaagaccgaat	1535
Q	1330	aagaatatctacacagacaaatgtagagaaagattatataccgaaatcaagagagaagc	1389
D	1536	aaaaacataatactgataatgaaagaaaaattcataatlgaaatgaagaaaaaat	1595
Q	1390	aaattgagaaagaaanaattgtagagttgacagaagaaagttaagaaagccgacgaaga	1449
D	1596	aaaaataaaaaaaaatttgatactgataaaaaacttaagaaagacagatcaagct	1655
Q	1450	ctaacagatcgaactaaagagtagaagaagctgtgaagagatcatgaatccaatc	1509
D	1656	ttaatgatataacaaaagaatgtagaaaaattcctaagaatttaagatgacaaatc	1715
Q	1510	aacaataacatcgactcgactacccaacttcgagaaatgatgtagaaaaagctactataca	1569
D	1716	aataataataatgatttaacttaattccgaaaaaatgtaggtaaaagatatcataaa	1775
Q	1570	gtgtagaagaactgtagacacccataacccttgcatccatggaattctaagcataatct	1629
D	1776	gttagaagaacttaacaccccttaacttgtgcctcctaagaattcttaacataatct	1835
Q	1630	gagagaagctacccaagagcttctaagtataagagagactatccctcggagaaactgtgtg	1689
D	1836	gaaagaagcttaacaaaagctcctaataatataggaagatatacttctaaggaatataagtt	1895
Q	1690	gagaaagaactaaagattatccaagaatctcatatagtaagatccgaaacgagatcgagcg	1749
D	1896	gaaagaagaatttaaaatatataaanaatttaataagaanaatagaaatgagattgaaca	1955
Q	1750	ctgttgtagaagaacttaagaagagatgtagaagaagctgtttggaagaaagattacaagaac	1809
D	1956	ttagttagaataatctaaaaaaagatgtagaagacagcttcttgtaaaaaaaattatacaagc	2015
Q	1810	gaaaataaacagagatgtagaagatcccgaggagctcccgatattgtttaagatccgaatgag	1869
D	2016	gaaataataacccagatgtaaaaaatttagaagatataccgcaattgttaaaagatacaattca	2075
Q	1870	aaggtgctcctcctatgaaaccaagatttgaactgaactaaagaaactcaactatctgaagac	1929
D	2076	aagttttatattatgaacaaaatttgcgcgaatttaaaaaagcctcaattgttttaaaaat	2135
Q	1930	gtgtagatttaaaacataataatactgtgcggaatagttaacgaggaagaaatgaagcagaa	1989
D	2136	gtagaatataaacaataataatactgtcccaactcttcaacaacaagaataaagaagaa	2195
Q	1990	ccatatactcctgtagtaactccaagaaagagatagacaaaactgaaggtgttaatgccaa	2049
D	2196	ccattattatttaattgtgtgtgaaaaaagaatttgataaatttaaaagtttcaatgctaa	2255
Q	2050	gtcgaagagctgtatcaacgtagagagaagaagaacatataaacgtgaagagcagctcagataac	2109
D	2256	gtagaatcatctgataaatagtagaaaaaaaatacaataaaaacaaagctcaatcgagataat	2315
Q	2110	tccgaagccttcacagaagagagataaacccggacaggtctacccaacaagcccggaacaag	2169
D	2316	tcggaacccatcaacgcgaagagaaataacaggaacaagcaactatacaaaactgtgacaaca	2375
Q	2170	gcgcgttcagctctcgaaagcgatagctgtgcgaagctcaagcacaagacagaagcagga	2229
D	2376	gcaggaatctgtctttagaagagagattcaagtaacagccaaagcacaagaacaaaacaaaga	2435

QY	2230	cagcttcagatgagccgaatgccccgttcagagaggtctaaagctccaagtctcaccacacagct	2289
Db	2436	caacacacagtaaccagtaaccagtaacgaagcaaaaagcaaaagctccacacacaccagca	2495
QY	2290	cctctgaataacaaagaccggaaatgtcagcaaaactgacataactctggagaagctcatgag	2349
Db	2496	ccagtaaatataaactgtgaaaatggttccaaattagattatctcttggaaaattatataca	2555
QY	2350	ttccgtgatatatcctacatctgcacaaatatatctctgtctctccacagcatatgac	2409
Db	2556	ttttaataacttaataatataatgcacaaatatatttggtttcacacacacatgagac	2615
QY	2410	gagaaattcttaacagtaacgataacaaagaaagagagatgaattgaactgccccgtgt	2469
Db	2616	gaaaagatatataaacaatatataaattacaaagagaggaagaaagcaaatgaattcatgt	2675
QY	2470	gaccacatgagcccgctgttcaatlatccagaaacacatcccggtatgatatgttc	2529
Db	2676	gaccatatagacttatgttttaataatacaaaataacataccctgtaattcatgttt	2735
QY	2530	gatggcctcaaca-----attct	2547
Db	2736	gatagtctaaacaatagttataccaactattatggtttatgaaaagaattggtat	2795
QY	2548	ctctctcaacagtttcatgagatataatgaaagagatgtgtcgaacatgataaactc	2607
Db	2796	gtatttatattacttatgtaaaatttatgaaaagaattggttctgtaattatataaact	2855
QY	2608	aaagaacaaacgaatataaagaacctcttcgaggaagctaaagaagctccacctctgt	2667
Db	2856	aagatatatgttcaaaattataaaatttatagagaaagcgaaaagaatgccacatctga	2915
QY	2668	aaaactctctcttcagctccacatgcacacacacgtctctccacaccccaagacaagccga	2727
Db	2916	aaaactcttcaagtctcaatgcacatgcacacattatcaactaacccctcgagataaaacggaa	2975
QY	2728	gtgagcgctaaacgacgacacctctcaactgcacacacttaactcaatgaacatggttt	2787
Db	2976	gtaagtgcgaatgtatgtatacatcatctcacaattttgaataatagttttaaattattt	3035
QY	2788	gagaaatcccggtctctccgcaagaataaagaacatcttacaagaacttatgtgaagaaga	2847
Db	3036	gaaaacatatgtagctctctgtgaaaaaacaataatatactaccaagaattataggtccaaaa	3095
QY	2848	tcgtccgagaacctctcaacgagaataactgaaagaacagcgacacatctataacgagagc	2907
Db	3096	agtagtgaaaacttttagaaagaatataaagaatagtagatactttataatgaact	3155
QY	2998	ttcaactactcgtggaattctaaagccgattgatatcaactctcttaacogataaactcaa	2967
Db	3156	tttcaaaatttctgtaaaatctaaagctgtgatatattcaatttcaatgaaatgaaacaaa	3215
QY	2968	cgtaagaagctggaagagggacatcaataagctgaaagaagacatgcaactgagcttgac	3027
Db	3216	aggaagaagaattagagaagagatatataataataaaaaaacttcagttatcatattgat	3275
QY	3028	ctgtacaaacagtacaaactgaaactctgagagagctcttcgacagaaagaagacagtcgcg	3087
Db	3276	ttatataataatataaattaaattagaagaatatttctgtataaaagaaaaacagttgct	3335
QY	3088	aagtataagatgacgatacaagaagtgtgactctgcgaaggagcagcttgaagaacaaact	3147
Db	3336	aaatataaaatgcaaatctaaaaaacttacttctataaaagaacaaatagaaatcaaaatg	3395
QY	3148	aaccctatgacaacatccgaaacacgctacgctgagaaactcttcagttctcttcaacaaag	3207
Db	3396	aattcacttaataaaccacaagcatgattatacaaaactttctgttcttcaacaaaaa	3455
QY	3208	aagaagaacgagatcgccgagacagaagaacactctgaggaacacacaagaattcttctcaa	3267
Db	3456	aaagaagctgaattatgagaagaacctgtgaaaacacattagaataacacaaaataattatgaaa	3515

OY	3268	cactacaagaagccttcgcaagttatataatgycgagttcttcctctcctcctcgaagctctcc	3327
OY	3268	cactacaagaagccttcgcaagttatataatgycgagttcttcctctcctcctcgaagctctcc	3327
Db	3516	catataaagagacttgtttaaattataatgttgaatcatccctaaataaacttaagt	3575
OY	3328	gaggagaagcatccagaccggggaataactcagccagctcgagaaacttaagttcgtct	3387
Db	3576	gaagatcaattcaaaaagaagaataattatgcagtttagaanaactttaaagttaaagt	3635
OY	3388	aagctcgaagcgaagctcgaagagacaacctcgaacctcggagaaagaagctcaagctcc	3447
Db	3636	aaattagaagaaataattaaaggatacttaattttagaanaagaaaaattatcatacta	3695
OY	3448	tctgcyggaactcgaactcaactcgtatccgcggagctcgaagagtcattaaagaagaactac	3507
Db	3696	tcaaggagttttacatcatttacttaatttgttcgatttaaaagaagtataaaaaataaattat	3755
OY	3508	accggcaatgcccagaacgggaataataacagcgtgaataacgcctcgtgaatttacaag	3567
Db	3756	acaggtatattctccaacgcgtaaaataacggagtgtaatacaatgcatctgaacttcaaa	3815
OY	3568	aagttcctgcctgaagaaacagaatgctgcgcacgttggttctcgtactcgtgcgcgaca	3627
Db	3816	aaattctccagaagaagaaacagatglttgcaacagtgtagaagtggaagtggaatccagaca	3875
OY	3628	cttgagcagctcacaactcaagaagcctcgtacatactacatcgtcggagccgagccaataca	3687
Db	3876	ttagaacaagaagtcacaaccaagaacacagatcaactcatcgttagagcgagagctcaaca	3935
OY	3688	attccaccatctcagaacgctcgcagcttgaggttcgaatgacagtcattatgttgctatctc	3747
Db	3936	atacaacaactcacaanaatgctcgaatgtagaagtagatgacgtatcatatgatactcaatt	3995
OY	3748	ggcgaagcgagagagagagctcagaatgacccgcgcgcgagtggtgtaacgcggagagctgaat	3807
Db	3996	ggagaatcccgaaagaagataatgtagtggtttagagcaagtgatgtaacggggaagcagtaact	4055
OY	3808	ccttcgctgtagataacaactctgtccaaaatcogaaacgaatacgaatgctcatctg	3867
Db	4056	aacttcgtaattgatacaactcttctaanaattgnaaatagtaatgagtttataattta	4115
OY	3868	aaacctctgacgagcgtctatagttctcccaagaacgcgtggaataataacgtgatgacc	3927
Db	4116	aaacctttagcagtggttttagaagtttaaanaaaacaaatgnaaataacgtatgaca	4175
OY	3928	ttcaatgtcaacgtgaagagcatctcgaacacgcggtttaatgaagagaagaataattcaag	3987
Db	4176	tttaatgttaatgttaagatataattttaaattccaagatttaataaacyggaanaatttcaa	4225
OY	3988	aagctcttgagagcgaactgtaattccctataaagaactgcctcctaactacgtgtgc	4047
Db	4236	aagttttagaatacagtttaacttcatatcctaataaagatttaacaatcaagtaattatgtgtc	4295
OY	4048	aaggaaccatacaagttctccatataaagaagaaggagataaatttcgttcaagtacaac	4107
Db	4296	aaagatccataataattcttctaataaagaanaaagagataaatttcaagaagttataat	4355
OY	4108	tatctoaagaccctcagccagccagatcatcaatttcgcgaatgagtgctggygattatc	4167
Db	4356	tattttaagatccatagatagatagcgaataaatttgcgaatgagtgctcgtgatatatt	4415
OY	4168	aagatcctgagcaaaaatacaagctcgtgacctgtaactatataaaaagtataacaagat	4227
Db	4416	aaaattatttcgaaaataataataacagtttagatttcaattataaaaaatatatacagac	4475
OY	4228	aaggaagcgagagatgaaaaatactcgcacctcctcgaataacatacgaaacctgtacaag	4287
Db	4476	aaaacaagtgtaaaatgagaaatacctcccttcttaaacaatattgagacctaataaa	4535
OY	4288	acagttgacacaaaatcgaccccttcgtgaattcaacctggaagggccaaggttcctaact	4347
Db	4536	acagtttaatataaatttgatttatttgaatttgaatttgaagagcaaaagttcctaattat	4595
OY	4348	actttagagagaagcgaattgtagagttaaaatcaaggaagctgaactacccaacaatc	4407

Db	4596	acatctggaatcaaacctgtagaagttaaataaagaacttaacttaataaacaatt	4655
Qy	4408	caagacaagctgagcagattccaagaaataaacaattctgcgtggaattgcagactgct	4467
Db	4656	caagacaatttgcagattttaaataaaaaaataaacatttcgttggaattgcgtgattatca	4715
Qy	4468	accgattataaccacaacaactccctcgacacaagttctctgcactgcatgtgttcgaa	4527
Db	4716	acagattataaccataataactattcttgacaagaattcttccttagtaacagatgtgtttgaa	4775
Qy	4528	aacctgcacaacaacagtgtctgagcaatctgtctoga - cggcaactgcagagcagtcgaa	4586
Db	4776	aactgtctaaatctcgtttattatcatattactgtgalttggaactgtgcagagtatgttaa	4835
Qy	4587	calctcccgacccaatgcgttggaagaacagtg - ccccaagaatagcgtgtttcagg	4644
Db	4836	catlccacaacaaccaatgcgtlaaaaaaacaatgataccacagagttctgtgattccaga	4895
Qy	4645	catctgagacgcgcgcgaaggtgcagaatgtctctcctgaaactaacaacaagaagagataag	4704
Db	4896	catttagtgaagaagaagagatgtlaaatgtttataatlaatacaacaagaaggtagttaa	4955
Qy	4705	tgcgttggaacccaacccctactctgcaatgaataacaatgctgggtgtgacgcgatgct	4764
Db	4956	tgtgtgaattcaatcaactctactcttgtaacgaaataatggtggtgtgacgtcagatgcc	5015
Qy	4765	aaatgcacgcgcgcgcgaagcagaacgcgtcttaacgcgaagaagaatacacatgctgagtgactaag	4824
Db	5016	aaatgtacccgaagaagatttcagtagtaacgcgaagaagaataacatctgtgaatgactaaa	5075
Qy	4825	ccgacactctatccacactcttcgcagcgatttttgcgcagctcacaatttcctctggcactc	4884
Db	5076	ctgatttgtatccactcttcgtatggtgaattttctgcgagttcctcctaactcttggaata	5135
Qy	4885	tccttctctgcatccatcctatgctgcatcctgttaacgcttcataatagatgcatg	4939
Db	5136	tcactctattataactcactgatttaataataatacagtttcatttlaaaaaatgttag	5190
RESULT 4			
AAQ80911			
ID	AAQ80911 standard; cDNA; 5181 BP.		
AC	AAQ80911:		
XX			
DT	24-AUG-1995 (first entry)		
XX			
DE	Plasmodium falciparum MSA-1 gene cDNA.		
XX			
KW	Plasmodium falciparum MSA-1 gene; recombinant poxvirus;		
KW	multicomponent multistage malarial vaccines; immunogens;		
KW	malaria diagnosis; ss.		
XX			
OS	Plasmodium falciparum (p486195).		
XX			
PN	WO9428930-A.		
XX			
PD	22-DEC-1994.		
XX			
PF	10-JUN-1994; 94MO-US06652.		
XX			
PR	11-JUN-1993; 93US-007583.		
XX			
PR	09-JUN-1994; 94US-0257073.		
XX			
PA	(VIRO-) VIROGENETICS CORP.		
XX			
PI	De Taisne C, Paoletti E, Tine JA;		
XX			
DR	WPI: 1995-036113/05.		
XX			
PT	Recombinant poxvirus contg. Plasmodium DNA in non-essential		
PR	region - useful in vaccines against malaria and for prodn. of		

PT Plasmidium immunogens
XX
PS Claim 3; Fig 6; 183pp; English.
XX
CC AAO80911 is the P. falciparum MSA-1 gene cDNA sequence. New
CC recombinant poxviruses containing either the SERA, ANRA, Pfmsp70,
CC AAM-1, Pfis25, Pfis16, CSP, PfisSP2, LSA-1, LSA-1 repeats, MSA-1,
CC MSA-1 (N-terminal p83 or C-terminal gp42) genes, or a combination
CC of these in non-essential regions of their genomes are claimed.
CC These poxviruses (pref. with a virulence reducing genomic
CC deletion or disruption) can be used as vaccines against malaria
CC and for the prodn. of Plasmidium immunogens. These viruses
CC provide multicomponent, multistage vaccines due to their expression
CC of sporozite, liver stage, blood stage and sexual stage proteins.
XX
SQ Sequence 5181 BP; 2287 A; 684 C; 707 G; 1503 T; 0 other;

Query Match 25.5%; Score 1262; DB 16; Length 5181;
Best Local Similarity 56.0%; Pred. No. 7.9e-297;
Matches 2910; Conservative 0; Mismatches 1995; Indels 291; Gaps 16;

QY 10 atgaataatcattctctccctgttcattctctgtttttatcaatcaatcagtcgtg 69
DB 1 atgaagatcattctctctgttcattctctgtttttatcaatcaatcagtcgtg 60
QY 70 accacagatcctatcagagctgttgaagaactggaagcttggagatgctcctt 129
DB 61 acacatgaagatcagagactgttcaaaaaacagagacttgaagatgagatg 120
QY 130 accgagatcagcctgttccagagaagagatggtcgtgaagaa----- 174
DB 121 acaggttctgtttttttcacaagaagaaatgacttaaatgaagaagaattacta 180
QY 175 -----ggagagatgagcagcgccgttacaacacagacacccgt 213
DB 181 aaagtgcagatgtcacaagtgtacaaagtgtacaaagtgtacaaagtgt 240
QY 214 tctaaaggtctgtgtcagcggtgtcgtgtgtgtgtgtgtgtgtgtgtgtgt 273
DB 241 acaagtgtacaaagtgtacaaagtgtcacaagtgtacaaagtgtcacaagt 300
QY 274 gcttcgagcagcagcgtgtcagtcagtcagtcagtcagtcagtcagtcagtc 333
DB 301 ggtacaagtgtacaaagtgtcacaagtgtgtacaaagtgtgtacaaagtgt 360
QY 334 cgaagaacaaatcactgtgaca----- 355
DB 361 agtccatcatctcgttcaacaacttacctcgttcaaatcttcaatctgtgt 420
QY 356 -----acttagagatccgagcgaagtcctcagcgcgacccaacacgg 408
DB 421 ccaagtgtcagaagcgttcaagatccttcaactcgtgtgttcaaaacagag 480
QY 409 aactatccccaactcaagaagcgtgaagtaaccacagctgtgtcagacctcaat 468
DB 481 aactactgttcacacttaagaactcaaatataccgaacctgtgtatcaacca 540
QY 469 atgtctgacactgtgtgataacatcattcgttcaaatatcgtattgacggt 528
DB 541 atgttaacttctgtgataacatcattcgttcaaatatcgtattgacggt 600
QY 529 atcaatgacccgtgtacaagtgtatttctactcgcagctgtgtgaagccaat 588
DB 601 atcaatgacatataataaacttaacttatttgaatttataagaagcaaat 660
QY 589 gacgtttgcgaactatgtcaaatccattcaattgaagaatgaagccaagag 648
DB 661 gatgatgtgtcaatgatattgtcaaatcattcaatccttaaatcgtgtgaag 720
QY 649 ttgagcgtatcgaagaagtgtgtctcgtgatacgaagcctcgcacaacaa 708
DB 721 tttagcgtacttaaaaaactgtgtgtgataatagaagaacattgaacatt 780

QY 709 aatgtggagaagtgtgaagattatataaaaaagaataagaagaccatcgaa 768
DB 781 aatgtgagaagaaatgtgaagatttactataaaaaataaaacacatcgaa 840
QY 769 gactgtacgaagaatccaaaaagcactagacaaaataagaatgtcacaaga 828
DB 841 gaattaatgtgaaggaagaaagaaacattgtcaaaataagaatgtcagata 900
QY 829 gaagaagaagtgtgtaccagccagtagcactgttccattataaaaaact 888
DB 901 ggaaaaaaaatltacaaagctcaataltatcttctatttaataaaacaa 960
QY 889 gaagcccaacctatcagctgtgtgagagcgcatagacacctcaagaaga 948
DB 961 gaagcacaattatataagcgttttgaaaaaactgtattgtcacctttaa 1020
QY 949 aatacaagaactgtcgaagaagattatgaatlaagaatccctcgcagcca 1008
DB 1021 aacttaaggaatttactgtgataagtaaatgaatlaaaatcccccacgg 1080
QY 1009 gggaaaccccttaacacgctgtcgtgacaagaacaaagatagagagcga 1068
DB 1081 ggaatcacccaataactctcctgtatagaacaaaaatcgaagacagaa 1140
QY 1069 atcaagaagatcgccaacacattgaagtccaagatgtcctccttccat 1128
DB 1141 ataaagaagaatgtcccaaaacatttaaatcattgtatgtatttaccat 1200
QY 1129 gactgtgactactgtgagagagaataagaatataagacatccgcgaag 1187
DB 1201 gaattagatatttatttaaggaaaaataaaagtgtgtatgaacccata 1260
QY 1188 -----gacaagaatcaacgacactatgaatatcccaatgtgtgtac 1239
DB 1261 gatccgaagaatcgttccaataatcaaaagtctccttccaaatggtat 1320
QY 1240 ctgtcttaacgaatataaaaa-----cgctccacagcgtcaatagct 1290
DB 1321 ttaccactcactgatatcaatcattcattcagctgtgagataagaataa 1380
QY 1291 gactgtatcaacccctcgtattatatacgaagaaccccttaagaatata 1350
DB 1381 gatttaatgaatccgtatcactaaagaataaataatgaataaattat 1440
QY 1351 gagagaagaagttatatacagaagaatcagaagatcaaatgtgaagaaga 1410
DB 1441 gaaagaataatattcattataacattaaaaacaaatgtgttgaagaaga 1500
QY 1411 gagagtgcagaagaagtgtacgaagccgcagcaaaagttcaaacgata 1470
DB 1501 aatcacacaaagaacaaaataaaatcttactgtgaagtttgaagaat 1560
QY 1471 tatgaagaagctgtgaaagatctatgtatccaatccaatcaatcagact 1530
DB 1561 tatgaagaatttactgtgaagaatlttatgaagaatlttaataatltt 1620
QY 1531 aacttcgagaagaatgtgaaacacgctactccttaaaatgtgagaacac 1590
DB 1621 gtctgtagaataatlttcaagtcagaatatacatatgtgtgaaaaacaa 1680
QY 1591 aataccttgcatacctatagagaatccttagacatccttggagagc 1650
DB 1681 aataaatttcaatccctcaaatcaatcgttatataatgttccaataat 1740
QY 1651 aaglatatgaggaactatctcgtcggaacattgtgtggaagaagaactaa 1710
DB 1741 tcatatcttgaagattatcttgaagaagaatlttctgnaaaagtlttaac 1800
QY 1711 aagaatcctaaatgaagaatcgaagaacgagatcgagacgctgtgtgaa 1770
DB 1801 tatcattgaaacactgctcgtgaagctgtatataaaaaataacagaaga 1860

OY	1771	gatgaagacagctgtgttggaagaagattacaanaagcgaataataacccagatvgagaag	1830
Db	1861	agtgaaacaaatctctgagaaaaaatcttlaaagcctaccactcagcaatg---ct	1917
OY	1831	atccggaaggtcccggtatgttlaaagccagatgicgagaagtgctccatgacaag	1890
Db	1918	tccttgaagatatagtatcttgaataatctacaagctcaaaaagttcttattaaaaaa	1977
OY	1891	atgtatgaactgaagaagactcaaccatctctgaaagaaagtgagttlaaacaata	1950
Db	1978	atagaagacttaagaagaagatagattatttttaaaaaatgcacaactaaagaatgaa	2037
OY	1951	catgtccgaatagtataagcagaagaataagcagaacactactactccatccgtactc	2010
Db	2038	catgtaccaatattttttlaaacacacaataaaccagaccatattatttaattgatt	2097
OY	2011	aagaagaagatagaacaacagcagaagtgctcatgcccagaagtgcagagcttataga	2070
Db	2098	aaaaaagaagtatgaataaataaaaagaatttatccaaaaagttaaaagacatgtaa--ga	2155
OY	2071	gagaagaagaacattlaaactctgaagacagctcagaataaactccagagcttccacaga	2130
Db	2156	aagacaacagctgtcttatacgaattataccacaaccccttagtgccagcaagcaaac	2215
OY	2131	gaagataaccggaacaggtctaccaccaagcccggaacaacagggccgtctcagagc	2190
Db	2216	aagat-ggggggtctactccacacacacattatcccaatccaggaacaagaataca	2274
OY	2191	gatagctgcgaagctcaagaacagagcagaagcagagccagcccggtgcagtgcc	2250
Db	2275	gaacacagaagaagaagaagaacagtagagacacacacaacaggtacaatacatlaca	2334
OY	2251	gttccagaagctlaaagctcagctgaagcttacaacacacagccctgtgataacaagac	2310
Db	2335	ccaaagaagtaaaagtgtgttgaaaattccaatagaacataagaagatagacaattaca	2394
OY	2311	aatgtccaaactggaactggaactccttgagaagctctatgagttccgtgaatacact	2370
Db	2395	gctcttgcaaaaacagttcttctaagaagaatttagatgaatttttaactaaactata	2454
OY	2371	tgcacacaatatatcccgctctccagcactatagaagaagagaattcttaaacagtac	2430
Db	2455	tgtcataatatatttttagaatcaactccttagatagagcaaaaattatagagtat	2514
OY	2431	aagaataacagaagagagaggaatgaactgctctctgtgataccagtgaagctgtcc	2490
Db	2515	aattctatccagaagaagagaatgaataaatacatgataccttagattattatt	2574
OY	2491	aatatccagaacaacatcccggtatgtatctctatgttcgatagcccaaatctctc	2550
Db	2575	aattataaatacatcactcgtctatgtattctatataatgatagatgaatacatgatt	2634
OY	2551	tctaaactgtctatgagatataatgagaagggatggtctgaacctgtataaactaaa	2610
Db	2635	caacatctcttttgaattataatcaaaaaggaatgattattatcatcaataactaaa	2694
OY	2611	gacaacgacagaattgaagaacctctcggaagagagctaaagaagctc-----	2657
Db	2695	gagggaataatcaatcaaaaaaattattagagggcaaaaaaataactgcgaatcatct	2754
OY	2658	-----caactcgttlaaactctcttc	2681
Db	2755	acatccagctcttgaaatacaaacgtaaatatctgtctcaatccgcgaactcacagtattcc	2814
OY	2682	cagttccatgcaaccactgtctctcaactcaagaca-----	2720
Db	2815	caaaaaccaacatcaaatgtactctctacaataccaaaatgtygtatgctgatact	2874
OY	2721	-----gcccggaattgagcgtctaacgacagacactcactgcaccaactcaatcaact	2775
Db	2875	ggtctcgtctgtatgtgaagaagatctatgctcccttaacagatattgtgattatgaagat	2934
OY	2776	ctgaacatgltttagaacaactcctgtctctcggcaagaataaagaact---ctacaagaa	2832

Db	2935	ttgaaaggtatgttgaatctcttcaaatcttgcgaataaacaataagtcacataccatta	2994
Qy	2833	cttatgtgacagaatctgcgcgaagctcttaagagaagatctcgaagacgcgacga	2892
Db	2995	accattctacacagagatgaggaataattttagaataattcttaaaaaataatgatacc	3054
Qy	2893	tctctaaacgagagctcactaaccttcgtgtaaatctcaaacgcgactgatatcaactctc	2952
Db	3055	tattttaatgtgtatcaacaacatctgcgtataaatctcaattcaaaagttaattcaagtttg	3114
Qy	2955	aacgataatctlaaacgttaagaagcttggaaagagacatcaataagctgaagagacactg	3012
Db	3115	acc---gaaacacaaaaaaatgcatataatgatagtgaattaaaaataaagaattacttta	3171
Qy	3013	caacgagcttcgcgcgcgtgacaacagtacaaactgaaactggaagactcttcgcagag	3072
Db	3172	cagttatcaatttgtattatataataaataaataaataaatgatagtattatctaataag	3231
Qy	3073	aagaagacagctgcgcgaagatlaaagatgacagatcaagaagattactctcgaagagcag	3132
Db	3232	aaaaaagacacttgcgcgaagcaaaaatgccaattaaaaaacttacttttaaaagaacaa	3291
Qy	3133	cttgaagcaaaactcaactcactgacaaatcggaaacacgatactgcgaactctcagtg	3192
Db	3292	ttagatctaaatattgaattcaacttaataaccacataatgatttcaaaaactttctgt	3351
Qy	3193	tctctcaacagaagaagagagccgagatcgcgcgagacagagaacaactcctgagaaccc	3252
Db	3352	tctcttcaacaaaaaaaagaagctgtgaattgagaaactcggaaacacatlagaaacaa	3411
Qy	3253	aagattctctccaaacactcaaaagcctcgtcaagtataataatgycgcgactctcctc	3312
Db	3412	aaaattattatgaaacattataaagaacttgtaaatattataatggtggaatcatctcca	3471
Qy	3313	ctgaagactctctccgagsgagacatccgaaccgagataactagacacgcagcctcgagaa	3372
Db	3472	ttaaaaactcttaagtgagatatacatccaacgagaataattatgccaattttagaaaaa	3531
Qy	3373	tccaaggtccgccttgaagctcgaagcgaagctggaaggaacactggaactcctggagaag	3432
Db	3532	ttagagattatlaagtaaaaatagatggaaaaactcaatgataattcacttggagaagaa	3591
Qy	3433	aagctcagctacccctctagcgcgacgtacatccactatcgcgcgactcaagaagactct	3492
Db	3592	aaattactcttctatacgaatgtgattatcaacgattatcgaattcaaaagaatata	3651
Qy	3493	aagaaacgaactacaccgcgcgaatagcccaagcggagaataatatacagacgtgaatacga	3552
Db	3652	aaaaataaaaaattatacagtaactctccaagtgaanaataaataagaagttacaagaact	3711
Qy	3553	ctggaatcttcaagaagcttcctgccttgaagaaacagatgctgcgaactggtgtctc---	3608
Db	3712	ttaaaactttagcgaaaattctctcccaagaacaaagtatacaacagttgtaactccact	3771
Qy	3609	-----tgatctgctcgcgcacacgcggcgagcttca	3641
Db	3772	caaccagatgtaactccatctccatctatctgttaaggtgaagtgtgagttcgaatccca	3831
Qy	3642	acctgaagaag-----cctgatactactcaltgtcggagccgagttccaatacatattac	3693
Db	3832	aaagaagaacaacataccaactctcagctcttattaaagaattacaacagtagta	3891
Qy	3694	acattctcgaagcgcgcgcgattgagctcgatagacgtcatatgtgtccatcttcggcgag	3753
Db	3892	caattacaaaattatgacgaagaagaatgtatcccttgattgattacccattcttggagaa	3951
Qy	3754	agcgaaggagactcagctgaccccgccgcgagtggtccacgcgtggaagctgcatacccttc	3813
Db	3952	tccgaagaataatgacgaataatttagatcaagtgaataactcggagaagc---aatatcgtc	4008
Qy	3814	gttatgtataaacattcgttccaaaactcgaaacgatacgaagtgccttatctgtaaacct	3873

Db 4009 acaatgataatctctcaggaattgaaatgaatatgatattataattaaacct 4068
 QY 3874 ctggcaggcgtctataggtctctcaagaagacgctgggaataagctgacgtccat 3933
 Db 4069 ttacgtgagatatacagaagcttcaaaaaacaattgaaaaaaacatttcaattat 4128
 QY 3934 gtcaacgtgaaggaatctctgacagcgctttaaagaagaataatttcaagacgtc 3993
 Db 4129 ttaatttgaacagatattcttaattcaacgtctcttaagaagaacgaataatttcttaagtta 4188
 QY 3994 ttggagagcagcttgatctccataaagaacctgacctctcttaactacgtgtgcaagac 4053
 Db 4189 ttgagatctgatttaattgaattcaattacatatatctcaatgaatatatttaagat 4248
 QY 4054 ccatacgaattctcaataaagaagaagggataatttctgtctagtctacacatatc 4113
 Db 4249 tcaatttaatttgaattctgacacaaaaacacacattttaaaggttacaatatata 4308
 QY 4114 aagagctcatcgacacagatatacaatttcgttaagtgtggtggtatcacagatc 4173
 Db 4309 aaagatcagtagaagaatagatatatttgcacaggaaggtatagttatataaag 4368
 QY 4174 ctgagcgaataatacagaactgacgtgactctatataaagaatatatacagaataag 4230
 Db 4369 gtttagcgaataataagatgatttagaatcaataaaagttatacaagaagaag 4428
 QY 4231 -----caa 4233
 Db 4429 gagaaatcccatcatcaccaacaacacctccgtcacacagtaaaaaacagaagaaca 4488
 QY 4234 ggcgagaaatgaataatctcctctctgaaataacatcgaaacctgtacagaacgtg 4293
 Db 4489 aagaaggaagtaagttctctccattttaaacaacatttgagacctatacacaattata 4548
 QY 4234 aagcgaacaaatcgacctctctgttaactcactgtgagcccaaggtctcactaactaac 4353
 Db 4549 gttataaattgacattacttaattacttaagaagcaatgaatgaattgattgt 4608
 QY 4354 gagaagagcaatgtgaggttaaaatcaagaagcgtgaactgaactaaacaatccaagac 4413
 Db 4609 gaataaggtgagcagcagtttaaaataactgaactagttgatttaaaagcaattgagac 4668
 QY 4414 aagctgagaaatltcaagaanaataacaaatttctgagaaatgacgacttctacgact 4473
 Db 4669 aaatagatctttaaacaacatacagactcgagacgaatlaaanaattgataatgat 4728
 QY 4474 tataaccacaacatctctgacacagtttctgtccactgacatggtgttcgaanaacctc 4533
 Db 4729 gatacgaanaaagatatactgtgcaaatctacttagtacagattag---ttcaaaattt 4785
 QY 4534 gccaaaacagtgctgagcaatctgtcgacgcaacctgcaagggacatgctgaacatctcc 4593
 Db 4786 cctaatacaataatacaaatlaattgaaggaanaattccaagataatgtttaaactttca 4845
 QY 4554 cagcaaccaatgcgtgaagaacagtgccccagaatagcggctgtttcaagcactgagac 4653
 Db 4846 caacacacatgcgttaaaaaaacaatgctcagaanaattctgtagtttcaagcatttagat 4905
 QY 4654 gagcggaagagtgcaagtgtctctgtaactacaacaagaagaaggaataagtggtgag 4713
 Db 4906 gaagaagagaatgttaaatgttttaattacaacaagaaggtgataatgtgtttaa 4965
 QY 4714 aaccacaacctacacctgaatgaanaacatgctgggtgtgacgcaatgaatgaaccc 4773
 Db 4966 aatccaactctactgttaacgaanaataatgtgtgagtgtgacgcaatgaaccc 5025
 QY 4774 gaggaagacagcgtctcaacaggaanaataacatgctgaggtgactaaagccgacatcc 4833
 Db 5026 gaagaagatcaggtgaggaacgaagaanaatacagatgtgactaaacactgatctc 5085
 QY 4834 tatcactcttcgacaggaatttttgcctcagcctcaatttccctggcactctcctctg 4893
 Db 5086 tatcactcttcgactgtatttctgcagttcctcctaactcttagaataatcattctta 5145

QY 4894 ctgactccatcgtatcctgtgacagcttcaactca 4929
 Db 5146 ttaatacctcagtgataatataacagttcatttaa 5181

RESULT 5

AAT97956
 ID AAT97956 standard; DNA; 1950 BP.

XX AAT97956;

DT 03-APR-1998 (first entry)

DE Chimeric MSA-1 antigenic protein 1 used in a malaria vaccine.

KW Signal peptide; malaria; vaccine; merozoite surface antigen-1 peptide;
 MSA-1; recombinant vaccinia virus; Plasmodium falciparum; anchor peptide;
 immune response; humoral; cell-mediated; merozoite; ss.

OS Chimeric - Mammalia.

OS Chimeric - Plasmodium falciparum.

OS Chimeric - Homo sapiens.

FH Key Location/Qualifiers
 FT 1..165
 FT misc-feature
 FT /tag= a
 FT /note= "signal sequence added to improve immunogenicity"

FT misc-feature

FT 166..1893
 FT /tag= b
 FT /note= "Plasmodium falciparum MSA-1 peptide"

FT misc-feature

FT 1894..1950
 FT /tag= c
 FT /note= "anchor sequence added to improve immunogenicity"

PN W09726911-A1.

PD 31-JUL-1997.

PF 29-JAN-1997; 97W0-US01395.

PR 29-JAN-1996; 96US-0593006.

PA (GEOU) UNIV GEORGETOWN.

PI Davidson EA, Yang S;

DR WPI: 1997-393372/36.

XX Malaria vaccine - comprises expression vector expressing fragment of
 PT merozoite surface antigen.

XX Claim 33; Fig 2; 75pp; English.
 PS

CC The present sequence encodes a chimeric protein that contains a
 CC mammalian signal and anchor sequence, and a merozoite surface
 CC antigen-1 (MSA-1) carboxy terminal peptide. The C-terminal fragment of
 CC MSA-1 provides a more specific response than the complete MSA-1,
 CC and attachment of anchor and signal sequences improve the immunogenicity
 CC of the protein better than the use of an adjuvant. The chimeric protein,
 CC and expression vectors (analogues that express MSA-1 without either
 CC signal or anchor peptides), particularly in the form of recombinant
 CC vaccinia virus, are used in vaccines to prevent or treat malaria caused
 CC by Plasmodium falciparum. The vaccinia vector expresses the antigen
 CC fragment for many days, or even years, generating a long-lasting immune
 CC response (humoral and/or cell-mediated) against the merozoite form of
 CC the parasite, in humans or other animals.

SO Sequence 1950 BP; 830 A; 236 C; 277 G; 607 T; 0 other;

Query Match 20.2%; Score 1000; DB 18; Length 1950;
 Best Local Similarity 72.4%; Pred.No.2.6e-233;
 Matches 1297; Conservative 0; Mismatches 495; Indels 0; Gaps 0;

QY 3138 aagcaactcaactctacgagaaacatccgaaacgctactgcgagacttctcagtttctt 3197
 DB 159 aaatgaattgaattactactaaactaaacgaagcatlatacaaaactttctgtttctt 218
 QY 3138 caacaagaagaaggaagcagatccgcgagacagagacacatccgagaaacccaagt 3257
 DB 219 taacaaaaaaagaagctgtaaatagcagaacgaaacacattagaacaaacaaaaat 278
 QY 3258 tcttcacaacatacaaaagcctcgtcaagatataatagcgagcttctccctctgaa 3317
 DB 279 attatggaacattataaagagcttgtaaatattataatgttgatccctccattaa 338
 QY 3318 gactctccgagagagacatccgacgagataactaagccagcctcgagaactcaa 3377
 DB 339 aacttlaagtgagaatcaatccaacagaaatattatgccaagtttagaanaactttaa 398
 QY 3378 ggtcctgtctaaagctcgagagcaagctgaagacaacctgaagcctggagaaagaagct 3437
 DB 399 agtattaagtaaatagaagaaatlaaaggataattaaattagaagaaaaaatt 458
 QY 3438 cagctacctctcctagcagctgcatcacctgactcgcgagctcaaggaagtcattaaaga 3497
 DB 459 atcatactatcaagtgattacatcatcttaatttgctgaattaaagaagtataaaaaa 518
 QY 3498 caagaactacacggcaatagccaagcgagataatacagaagctgaataacgcacttga 3557
 DB 519 taaaaattatcaggtlaattctcccaagtgaaataatacagtgtaacaatgacttaga 578
 QY 3558 attcttaagaaggtccgcctcgagagacagatgtgcgccactgtggtgtcgaatcgg 3617
 DB 579 atcttacaacaaatctcccccagaagacagatgtgcacaagctgtgaaggaatgg 638
 QY 3618 ctccgacacactgagcagctctcaacctaaagagcctcatctactatgtctcgagcgca 3677
 DB 639 atccgacacattagacaagaagtcacaacaaagaaacacatcatcattgttggcgca 698
 QY 3678 gtccaatacaattaccacatctcagaacgtcgagatgagtgactgaacgtcatcatgtt 3737
 DB 699 gtctacacaaatacaacaatcacaaatgtcgatgataagtagtgacgtatcatagtt 758
 QY 3738 ggcctatctggcgagagcgagagactagactcctgcgcagcggtgtacacgggca 3797
 DB 759 acctatatttggagatcccgagaagattatgattagagcaagtagtaacgagaga 818
 QY 3798 ggcctgcacactcctcgatgtgataactctgtccaaatcgagaacgaaatcagaagt 3857
 DB 819 agcagtaactccttcgcgaattgtataacatacttcttaaatltgaatgaatgaaggt 878
 QY 3858 gctctatctgaacactctgcgagcgctctatagctctcacaagaacagctlgagaataa 3917
 DB 879 ttatatttaaaacctttagcaggtgttlatagaaagtttaaaaaacatttagaanaataa 938
 QY 3918 cgtgatgaccttcaatgctcaacgctgaagacatcttgaacagcgcttataagaaga 3977
 DB 939 cgttatgacatttaagttaagttaagatatttaaatltaacgatttataaagcgtga 998
 QY 3978 aaattcaagaacgcttgagagagcagcttgattccctataagaacccgagactctctaa 4037
 DB 999 aaattcaaaaatgttttagaatcagatttaattccatataagaatttaacatcaagtaa 1058
 QY 4038 ctacgttgcagaagccatacaagttcctcaataaagaagaaggaataaattctgtc 4097
 DB 1059 ttatgttgcagaagccatataaattcttaataagaagaaagagataaattcttaag 1118
 QY 4098 tagttacaacatatacaagagactcatcgacacgcatatcaattcgcgttaagatgtct 4157
 DB 1119 cagttataattatlaaggaattcaatagatagcagataaatttgcgaatgagtgctt 1178
 QY 4158 ggggtattacaagaatcccgagcgaaataatacaagctgaccttgactctattaagaagta 4217

DB 1179 tggattatataaataatataatccgaaaaataataacagattagatttcaatlaaaaaata 1238
 QY 4218 tatcaacgataagcagcgagaaatgaaaaataatctgccttccctcgataaatatcgaaac 4277
 DB 1239 tatcaacgacaaacaaagtgaaaaatgaaataacatccctctttaaacaatattgagac 1298
 QY 4278 cctgtacaagaagcgtgaacgacaaatcgacctcttgtaacttcaaccgagagccaagt 4337
 DB 1299 ctatataaacaagttaatgataaaatgtatatttattgtaattcattagaagcaaaagt 1358
 QY 4338 cctcaactatacttcaagaaagcagatgtggaagttaaatacaagagagcgtgaactact 4397
 DB 1359 tctaatattatacatatgagaaatcaaacgtagaagttaaataaagaacttaattactt 1418
 QY 4398 caaaacatccagaagcagctgacagatttcaagaaaaataacaattcgttcggaattgc 4457
 DB 1419 aaaaacaattcaagaacaaatlgcagattttaaanaaaataacaattcgttgcgaattgc 1478
 QY 4458 agacctgtctaccgattataaccacaacatctcctgaccaaagtttctgtccactgcat 4517
 DB 1479 tgattatcaacagattataacataataactattgacaagatcccttagtacaggtat 1538
 QY 4518 ggtgttcgaaaaaccccgcaaaacagtgctgagcaatcgtctcgacgagcaacctgcaggg 4577
 DB 1539 ggttttgaanaactctgtcaaaacccgttttacttaattacttgtagaacttgcaggg 1598
 QY 4578 catgctgaacatctcccgacaccaaagtcgttagaagaacagatgtcccccagaatagcgctg 4637
 DB 1599 tatgttaaacatttcaacaccaaagtcgtaaaaaacaatgtccacaaattctcgatg 1658
 QY 4638 ttctgagatctgcgagcgcgagagatgtaagtgctcctcgaactcaacaagaagg 4697
 DB 1659 ttccagacatttagatlgagaaggaagatgtaaaagttatattaaatacaacaagaagg 1718
 QY 4698 agataagtgctgagaaaccccaacccctacatgcaatgaaacaaatctgcgggtgagac 4757
 DB 1719 tgataaatgtgttgaaatcccaatccctactgttagaagaaataatgttggatgagtc 1778
 QY 4758 cgaatgtaaatgacagaggaagacagcggtctaaacggaagaaatcatcatgagtg 4817
 DB 1779 agatgcacaattagcagaagaagatcagtagcaacggaagaaatcatcatgagtg 1838
 QY 4818 tactaagccgactctctccactcttcgacggagttttttgtccactcaatttct 4877
 DB 1839 tactaaactgtacttcaacacttctcgatgttatttctgcagtlcctcctaactctt 1898
 QY 4878 gggcatcctcctcgtcgatcctcatgtgactgtgactgtacagcttcaatcaa 4929
 DB 1899 aggaatcatcttataactataactcaatgttaatatataacagttcatttaa 1950

RESULT 6
 AAT97958
 ID AAT97958 standard; DNA: 1897 BP.
 AC AAT97958;
 DT 03-APR-1998 (first entry)
 XX Chimeric MSA-1 antigenic protein 3 used in a malaria vaccine.
 DE
 XX
 XX Chimeric MSA-1 antigenic protein 3 used in a malaria vaccine.
 KW Signal peptide; malaria; vaccine; merozoite surface antigen-1 peptide;
 MW MSA-1; recombinant vaccinia virus; Plasmodium falciparum; anchor peptide;
 KW immune response; humoral; cell-mediated; merozoite; ss.
 OS Chimeric - Plasmodium falciparum.
 XX Chimeric - Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT misc_feature 1..1839
 FT /tag= a
 FT /note= "Plasmodium falciparum MSA-1 peptide"

Dh 421 atcaacgaataagcaggaagacgaagtaactctgctccctctcgtgaacaacatcgagacc 480
Oy 4279 ctgtacagaagcagtgacacaaatcgacctcttcgttaattcaccctggaggccaagctc 4338
Dh 481 ctgtacaagaacggtacgaataagatgtctctgtctgtaaccctggaggccaagctc 540
Oy 4339 ctcaactaacttaagaaagcaatgttggaagttaaatcaagagactggaactacac 4398
Dh 541 ctgaactacactatgagaagaacgctgaggtgcaagatcaagagctgattacactg 600
Oy 4339 aaacaatccaagaacagctgagcaatctcaagaanaataacaattcgtcgaaatgca 4458
Dh 601 aagaccatccaggaataagctgcccgtatccaagaagaacaacaactcgtcggaatcgcc 660
Oy 4459 gaccgtctaccgattataaccacacaactctcctgaccgaattctctccactgactg 4518
Dh 661 gactcgagaccggaattacaacacacaacactcgtcgacaagaattccttgagaccggtatg 720
Oy 4519 gtgttcgaaaacctcgccaacaaacagctgtgagcaatctgctcgagcaacactcgagggc 4578
Dh 721 gtcttcgaaaacctcgccaacagcgtctcctgagcaacactcgtgagatggagactcgagggg 780
Oy 4579 atgctgaacatctccagcaccacatcgctgaaagaacagctgccccagaataagcggtgt 4638
Dh 781 atgctgaacatcagccagcaccagctgtgaaagcagctgtcccccagaacagcggtgt 840
Oy 4639 tttaggcactcgagcagcgggaagatgtaagctgtctcctggaactacaacaaagaaga 4698
Dh 841 ttcaaacaccctggaatgagagagagatgtaagctgtcgtgtaactacaagaagaaggt 900
Oy 4699 gataagtcgtgagagacccaacacctctacactcgtcaatgaaacacagcggtgtgagacc 4758
Dh 901 gataagtcgtgagaaacccaacacctctactctgtaacgagaacacagtggtgagatgacc 960
Oy 4759 gatgctaaatgcacgaggaagacagcggtctcaacggaagaataacacatcgagatgt 4818
Dh 961 gatgcaagctgtacagagagatctcaagggagcaggaagaagatcacctgtgagatgt 1020
Oy 4819 actaagccgactcctacacactcctcgacgggagattttg 4859
Dh 1021 accaagcctgattctatctacacactgttcgattgagatctctg 1061

RESULT 9
AA25586
ID AAX25586 standard; CDNA: 1065 BP.
AC AAX25586;
XX 02-AUG-1999 (first entry)
DT
XX Merozoite surface protein MSP-1-42 modified cDNA.
DE
XX MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
transgenic animal; ss.
XX Plasmodium falciparum.
OS Synthetic.
XX
PN WO9920766-A2.
XX 29-APR-1999.
PD
XX 20-OCT-1998; 98WO-US22225.
PF
XX 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
XX (GENZ) GENZYME TRANSGENICS CORP.
PA
XX Chen LH, Meade H;
PI

DR WPI: 1999-302742/25.
DR P-PSDB: AAY05832.
XX
PT New modified recombinant nucleic acid sequences useful for producing
PT malarial DNA vaccine
XX
XX
PS Claim 12; Fig 1; 43pp; English.
CC This novel, modified nucleic acid encodes a 42 kDa C-terminal
CC part (see AAY05832) of malarial merozoite surface protein MSP-1
CC (MSP-1-42), an important target for the development of a vaccine
CC against Plasmodium falciparum. The nucleic acid sequence has been
CC modified compared to the native sequence of MSP-1-42 (see AAX25587)
CC such that 306 nucleotide positions have been replaced to lower the
CC AT content (from 76 to 49.7%) and to eliminate 10 mRNA instability
CC motifs while maintaining the same protein amino acid sequence.
CC These alterations allow MSP-1-42 to be expressed in mammalian cell
CC culture and in transgenic mice. The invention provides modified
CC recombinant nucleic acid sequences and methods for increasing the
CC mRNA levels and protein expression of proteins that are difficult
CC to express in cell culture systems, mammalian cell culture systems
CC or in transgenic animals. The preferred difficult protein
CC candidates for expression are those derived from lower organisms
CC such as parasites, bacteria and viruses that have DNA coding
CC sequences of high AT content or which have mRNA instability motifs
CC or rare codons relative to the recombinant expression system to be
CC used. The invention allows expression of MSP-1 protein in the milk
CC of transgenic animals, and also provides a DNA vaccine comprising a
CC vector containing the altered MSP-1-42 sequence.
XX
SQ Sequence 1065 BP; 324 A; 254 C; 280 G; 207 T; 0 other;

Query Match 15.1%; Score 744.2; DB 20; Length 1065;
Best Local Similarity 81.3%; Pred. No. 3.3e-171;
Matches 863; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

Oy 3799 gctgcaactccttcgtgatataacattctgccaatctggaagaagaagtg 3858
Dh 1 gctgcaactccttcgtgatataacattctgccaatctggaagaagaagtg 60
Oy 3859 ctctactgaacactctgcaagcgctctatagctctcaagaagaacgctggagaataac 3918
Dh 61 ctgtacactgaagcgcgtgcaaggggtctaacccggagcctgaagaagcagctggagaataac 120
Oy 3919 gtgatgaccttaatgtcaacgtggaagacatctcgaaacgcgcttaataagaagaaga 3978
Dh 121 gtgatgaccttaacgtggaagacatctcgaaacgcggttcaacaacgcggagag 180
Oy 3979 aattcaagaagctgttgagaggaactgtattcctcttaagaagacctccctctaac 4038
Dh 181 aacttcaagaagctgttgagaggaactgtattcctcttaagaagacctccctctaac 240
Oy 4039 taagtgtcaagaagccctatacaagttcctcaataaagaagaaggaataaattctgct 4098
Dh 241 taagtgtcaagaagccctatacaagttcctcaagaagaaggaataaattctgctg 300
Oy 4099 agttacaactatatacaagactcgcgaacacgatatcaatttcgtcaatgagtgtctg 4158
Dh 301 agttacaactatatacaagactcgcgaacacgatatcaatttcgtcaatgagtgtctg 360
Oy 4159 gggattacaagaactcgtgagaagaataatcaagctgacactgtgattataaagaat 4218
Dh 361 gggattacaagaactcgtgagaagaataatcaagctgacactgtgattataaagaat 420
Oy 4219 atcaacgaataagcaggaagatgaataatactgacctcttccttgaataacatcgaaac 4278
Dh 421 atcaacgaataagcaggaagatgaataatactgacctcttccttgaataacatcgaaac 480
Oy 4279 ctgtacaagaacgtgaagacacaaatcgacctcttcgttaattcaccctggaggccaagctc 4338
Dh 481 ctgtacaagaacgtgaagacacaaatcgacctcttcgttaattcaccctggaggccaagctc 540

Claim 12; Fig 11; 43pp; English.

This novel, modified nucleic acid encodes a 42 kDa C-terminal part (see AAY05894) of malaria merozoite surface protein MSP-1 (MSP-1-42), an important target for the development of a vaccine against Plasmodium falciparum. The nucleic acid sequence has been modified compared to the native sequence of MSP-1-42 (see AA25587) such that the AT content has been reduced and 10 mRNA instability motifs eliminated while maintaining the protein amino acid sequence. In addition, a sequence encoding a 15-amino acid beta-casein signal peptide has been added to the 5' end of the sequence, and N262Q and N181Q mutations have been introduced to eliminate N-glycosylation sites. These alterations allow MSP-1-42 to be expressed in the mammary gland (i.e. milk) of transgenic mice. The invention provides modified recombinant nucleic acid sequences and methods for increasing the mRNA levels and protein expression of proteins that are difficult to express in cell culture systems, mammalian cell culture systems or in transgenic animals. The preferred difficult protein candidates for expression are those derived from lower organisms such as parasites, bacteria and viruses that have DNA coding sequences of high AT content or which have mRNA instability motifs or rare codons relative to the recombinant expression system to be used. The invention allows expression of MSP-1 in the milk of transgenic animals, and also provides a DNA vaccine comprising a vector containing the altered MSP-1-42 sequence.

Sequence 1140 BP; 353 A; 282 C; 290 G; 215 T; 0 other;

XX
XX
DT
Nov 1954

OY 4331 ccaaggtcctcaactatacttacgagaagagcaatgtggaagttaaaatcaagagctga 4390

```

Db      578 ccaagtcctgcagtagacacatagagaagcaacgtgagtgcaagttcaagagcgcga 637
      ||||||| | | | | ||||||| | | | | ||||||| | | | | |||||||
Qy      4391 actacctaaacaatccaagaacgctgcagatattcaaaaaatacaatttcgtcg 4450
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      638 attaccgaacacatccaagaataagctggccgattcaagaagacacacattcgtcg 697
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Qy      4451 gaattgcagacctctccgattacaaccaaatctctccgcaagttctgtcca 4510
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      698 gaatgcagacctctccgattacaaccaaatctctccgcaagttctgtcca 757
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Qy      4511 ctgcgaatggtctcgaaaacctcgccaaaacagtgctgagcaatctgtcgagcggaac 4570
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      758 ccgaatgctcttcgaaaaacctcgccaaagccgtctctgagcaacctgtgtatgtgaac 817
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Qy      4571 tgcagggagatgtgacatctcccgaccgaatgctgtgaagaagaagtgccccagata 4630
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      818 tgcagggagatgtgacatctcccgaccgaatgctgtgaagaagaagtgccccagata 877
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Qy      4631 gcgcgtgttcaggacatctgagcagcgagagagtgcaagtgctctctggaactacaac 4690
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      878 gcgagatgcttcagacacctgtgagagagagagagtgcaagtgctctggaactacaac 937
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Qy      4691 aagaagagagaaagtgctgtgagagagagagagagagagagagagagagagagagagag 4750
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      938 aggaagagagaaagtgctgtgagagagagagagagagagagagagagagagagagag 997
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Qy      4751 gtgcagcgcgaatgcaatagcccgagagagagagagagagagagagagagagagag 4810
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      998 gcgagtgcttcagacacctgtgagagagagagagagagagagagagagagagagag 1057
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Qy      4811 gcgcgtgttcagacacctgtgagagagagagagagagagagagagagagagagag 4860
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      1058 gcgcgtgttcagacacctgtgagagagagagagagagagagagagagagagagag 1107
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||

```

RESULT 12

AAC68976
ID AAC68976 standard; DNA; 1077 BP.

XX AAC68976;

XX 27-FEB-2001 (first entry)

XX Merozoite surface protein-142 coding sequence.

XX Merozoite surface protein; protazoacide; vaccine; malaria; ss.

XX Plasmodium falciparum.

XX WO200063245-A2.

XX 26-OCT-2000.

XX 20-APR-2000; 2000WO-GB01558.

XX 20-APR-1999; 99GB-0009072.

XX 13-MAY-1999; 99US-0311817.

XX 25-MAY-1999; 99CA-2271451.

XX (MED-) MEDICAL RES COUNCIL.

XX Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaiybull C;

XX WPI; 2001-015762/02.

XX Novel variants of the C-terminal fragment of Plasmodium merozoite

XX surface protein-1, useful as vaccines for treating or preventing

XX malaria -

XX Example 5; Fig 15; 126pp; English.

XX The present invention relates to non-natural variants of a C-terminal

CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC non-natural variants have reduced affinity for at least 1 antibody
CC capable of blocking a second antibody that inhibits the proteolytic
CC cleavage of Plasmodium MSP-1₄₋₂, and has the same affinity for at least
CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC MSP-1₄₋₂, compared to natural MSP-1₁₋₉. The non-natural variants of the
CC present invention are useful for immunising a mammal against malaria, and
CC can be used to treat malaria. The present sequence is the MSP-142
CC coding sequence.

XX Sequence 1077 BP; 338 A; 228 C; 225 G; 286 T; 0 other;

Query Match: 14.6%; Score 719.8; DB 22; Length 1077.
Best Local Similarity 79.3%; Pred. No. 2.9e-165;
Matches 853; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

```

Qy      3799 gctgtaacctcttcgtgattgataacatctgtccaaaacggaagaaatagagtg 3858
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      1 gctgttaacctcttcgtgattgataacatctgtcttaagattggaagaaatagagtg 60
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Qy      3859 ctctatctgaacctctgagagcgctctatagttctctcaagaacagctgagaaataac 3918
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      61 ttgtacttgaagcctctgcccgtgtctacagatccctgaagagcaactlgaaaaacaac 120
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Qy      3919 gtgataaccttcaatgtcaagtgagagacatctctgaacagccgcttataagaagagaa 3978
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      121 gtatgatttcaacagtttaacgttaacagagacatcttgaaacttcaacaaagagaa 180
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Qy      3979 aattcaagaagctcttggagagcagctgtatctccataaagaagcctgacctcttaac 4038
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      181 aacttcaagaagctcttggagagcagctgtatctccataaagaagcctgacctcttaac 240
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Qy      4039 taagtgtgcaagagcctataaagttctcctaataagaagagagagaaattctgtct 4098
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      241 taagtgtgcaagagcctataaagttctcctaataagaagagagagaaattctgtct 300
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Qy      4099 agttacaactatcaagagacttccatcgacacgacatcaatctcgtaagatgtgtg 4158
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      301 tcttaacaactatcaagagacttccatcgacacgacatcaatctcgtaagatgtgtg 360
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Qy      4159 gggattataagaagctctgagcgaataatatacaagcttgaccttgactctataaagat 4218
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      361 ggttactataagaagctctgagcgaataatatacaagcttgaccttgactctataaagat 420
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Qy      4219 atcaagataagaagcgaagagaaatataatctgacctctcgtaataatagaac 4278
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      421 atcaatgataagaagcgaagagaaatataatctgacctctcgtaataatagaac 480
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Qy      4279 ctgtacaagaagctgagcgaacaaatcgacctctgtaattcaacctggaagcgaagct 4338
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      481 ttgtacaagaagctgtaagataagaatcgatctgtgtcattcaatttggaaagcaagct 540
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Qy      4339 ctcaactatacttcaagaagaagcaatgtggaagttlaaatacaagagcgtgaactctc 4398
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      541 ttgcaatacactttagaagaagctgacgtgaggtcaagatcaaggaattgaactctc 600
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Qy      4399 aaaaacaatccaagaagcgtgagcagatttcaagaataatacaattcgtctggaattgca 4458
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      601 aagaacatccaagaagcgtgagcagatttcaagaataatacaattcgtctggaattgca 660
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Qy      4459 gactgtctacgatttataaccacaacatctcctggaacagttctgacacgagatg 4518
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      661 gattgtccactgtattacaacacaacatctgtgactaagttccgtctaccggaatg 720
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Qy      4519 gtttgcgaaaaacctcgccaaaacagtgctgagcaatctgctcgacggcgaactlgaaggc 4578
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      721 gtttgcgaacttgcgaagagcgtctctgacacacgttgatgtgtaactcagaggt 780
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Qy      4579 atgtcgaacatctccagcagcaaatgtcgtagaagaacagtgccccagaaatagcgctgt 4638
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||
Db      781 atgtgacacattgccacaacacaaatgtctgaagaagaatgtcacacaacactccgagatg 840
      ||||| | | | | | ||||| | | | | | ||||| | | | | | |||||

```

OY 4639 ttcaggcatctggacgagcggaagtgcaagtgctctctgaactacaagaagga 4698
 Db 841 ttcagacacttcgacgagagaagaatgtaagtgctgttgtaactacaagaagga 900
 OY 4659 gataagtcgctggagaacccaacctctacgcaatgaaacaatggcggtgtgagcc 4758
 Db 901 gataagtcgttgagaacccaacctctacgcaatgaaacaatggcggtgtgagcc 960
 OY 4759 gatgctaattgacgagagacgagcgctcttaacggaagaataacatgagagtg 4818
 Db 961 gatgctaattgacgagagacgagcgctcttaacggaagaataacatgagagtg 1020
 OY 4819 actaagcccgacccatccatccactcttcgacgagatcttcttcgaagctcaat 4873
 Db 1021 actaagcccgacccatccatccactcttcttcgagatcttcttcgaagctcaat 1075

 RESULT 13
 AAX56009 standard; CDNA: 1088 BP.
 AAX56009;
 AAX56009;
 31-AUG-1999 (first entry)
 Merozoite surface protein MSP-1-42 CDNA.
 MSP-1; merozoite surface protein; malaria; vaccine;
 protein engineering; protein expression; codon usage;
 transgenic animal; ss.
 Plasmodium falciparum.
 Key Location/Qualifiers
 FT 1..1086
 FT CDS /*tag= a
 W09920774-A2.
 29-APR-1999.
 20-OCT-1998; 98WO-US222226.
 15-MAY-1998; 98US-0085649.
 20-OCT-1997; 97US-0062592.
 (GENZ) GENZYME TRANSGENICS CORP.
 Chen LH, Meade H;
 MPI: 1999-288313/24.
 P-PSDB; AAY09373.
 Modified malarial protein for use in anti-malarial vaccines
 Example; Fig 2; 35pp; English.
 This nucleic acid encodes a 42 kDa C-terminal portion (see AAY09373)
 of malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
 important target for the development of a vaccine against
 Plasmodium falciparum. The 3' end of the sequence has been
 modified to include a 6xHis tag. The nucleic acid sequence has
 been modified (see AAX56008) according to a method of the invention
 in order to improve expression in mammalian host cells and in
 transgenic animals. In the modified coding sequence, 306 nucleotide
 positions have been replaced to lower the AT content (from 76 to
 49.7%) and to eliminate 10 mRNA instability motifs. The encoded
 amino acid sequence is unaltered. In another modified sequence
 (see AAX56021), a signal peptide sequence has been added and two
 N-glycosylation sites eliminated. The invention allows expression
 of MSP-1-42 protein in the milk of transgenic animals, and also
 provides a DNA vaccine comprising a vector containing the altered
 MSP-1-42 sequence.

XX SQ Sequence 1088 BP: 454 A; 135 C; 150 G; 345 T; 0 other:
 Query Match: 12.7%; Score 625.2; DB 20; Length 1088;
 Best Local Similarity 74.3%; Pred. No. 3e-142;
 Matches 789; Conservative 0; Mismatches 273; Indels 0; Gaps 0;
 OY 3799 gctgtcactccttcgctgattgatacttctgtccaaatcgtagaagaatgagtg 3858
 Db 1 gacgaactcccttcgctgattgatacttctgtccaaatcgtagaagaatgagtg 60
 OY 3859 cctctctgaaccccttcgagcgctcctataagtgctcccaagaacgctgagatc 3918
 Db 61 ttatttttaaaccttcgagcggtgtttatgaaatttaaaaaaatatgaataac 120
 OY 3919 gtgatgaccttcaatgtcaacgtgaagacatcttgaacgcgctttaaagaagaa 3978
 Db 121 gtatgacatttaattgtaattgttaagatattttaaattcagatttaatacgtgaa 180
 OY 3979 aattcaagaagctcttggaagcgacttgatccctataaagacctgacctcttac 4038
 Db 181 aattcaaaaaatgttttagaactcagatttaattcctaataaagatttaacaagtat 240
 OY 4039 taagtgtcaaggaaccatacagaatcctcctaataagaagaaggagataattctgct 4098
 Db 241 tatgtgtcaagaagatccatataatttcttaataagaagaagaagaataattcttaagc 300
 OY 4099 agttacaactatacaagaactcctacgcacacgataatcaatttcgtaattgtgctg 4158
 Db 301 agttataattataaagattcagatcagatacagataaatttgcgaattgtgctc 360
 OY 4159 gggatttacaagaatcctgaagcaaaaatacaagcttgacgtgactctatttaagaat 4218
 Db 361 ggaattataaataattatccgcaaaaatacaataagatttgaattcaataaataat 420
 OY 4219 atcaacgataagcaagcgagagataaataatactgccttcctgaataacatcgaacc 4278
 Db 421 atcaacgataagcaagcgagagataaataatactgccttcctgaataacatcgaacc 480
 OY 4279 ctgtacaagacgtgaacgacaaaatcgcactcttcgtaattcaccttgaagcgcaagtc 4338
 Db 481 ttataataaacgattgaataaattgattatttgaattcattgaagaacaaagtt 540
 OY 4339 ctcaactactactagaagaagcaaatgttggaagttaaatcaagaagctgaacactc 4398
 Db 541 ctcaattatcatcagaaatcaaacgtagaagttaaatcaagaagctgaactacta 600
 OY 4399 aaaaacatccaagaagaagctgacgaatttcaagaataaatacaatttcgtcggaattga 4458
 Db 601 aaaaacatccaagaagaagctgacgaatttcaagaataaatacaatttcgtcggaattga 660
 OY 4459 gaccgtctcagattataaacacacaaatctcgcagcaagttcttcctcgtgacatg 4518
 Db 661 gattatcaacagattataaacacataaactatgtgacaagaagttccttagacagatg 720
 OY 4519 gtgttcgaaaacctcgcacaaaacagtgctgagcaatctgctgcagcgcaacctgagggc 4578
 Db 721 gtgttcgaaaacctcgcacaaaacagtgctgagcaatctgctgcagcgcaacctgagggc 780
 OY 4579 atgcgacatctccagacacaaatgctgaaagaagaagtgccccgaataagcgcgctg 4638
 Db 781 atgttaaacattccacacacaaatgctgaaagaagaagtgccccgaataatctgagatg 840
 OY 4639 ttcaggcatctggacgagcggaagtgcaagtgctctctgaactacaagaagga 4698
 Db 841 ttcagacacttcgacgagagaagaatgtaagtgctgttgtaactacaagaagga 900
 OY 4699 gataagtcgctggagaacccaacctctacgcaatgaaacaatggcggtgtgagcc 4758
 Db 901 gataagtcgttgagaacccaacctctacgcaatgaaacaatggcggtgtgagcc 960
 OY 4759 gatgctaattgacgagagacgagcgctcttaacggaagaataacatgagagtg 4818

Db 961 gatgcgaatgacgcgaagaagattcaggtagcgaacggaagaataatcaatgtaagt 1020
OY 4819 actaagcccgactcctatccactcttcgacggatttttcg 4860
Db 1021 actaaacctgattcttaccactcttcgatggtatttttcg 1062

RESULT 14
AAK25587
ID AAK25587 standard; CDNA: 1088 BP.
XX
XX AAK25587;
XX
DT 02-AUG-1999 (first entry)
XX
DE Merozoite surface protein MSP-1-42 cDNA.
XX
XX MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KM transgenic animal; ss.
XX
OS Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT CDS 1..1086
FT /*tag= a
XX
XX MO9920766-A2.
XX
XX 29-APR-1999.
XX
XX 20-OCT-1998; 98WO-US22225.
XX
XX 15-MAY-1998; 98US-0085649.
XX 20-OCT-1997; 97US-0062592.
XX
XX (GENZ) GENZYME TRANSGENICS CORP.
XX
XX Chen LH, Meade H;
XX
XX WPI; 1999-302742/25.
XX
XX P-PSDB; AAY05833.
XX
XX
XX New modified recombinant nucleic acid sequences useful for producing
XX malarial DNA vaccine
XX
XX
XX Disclosure; Fig 2; 43pp; English.
XX
XX This nucleic acid encodes a 42 kDa C-terminal portion (see AAY05833)
XX of malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
XX important target for the development of a vaccine against
XX Plasmodium falciparum. The 3' end of the sequence has been
XX modified to include a 6xHis tag. The nucleic acid sequence has
XX been modified (see AAK25586) according to a method of the invention
XX in order to improve expression in mammalian host cells and in
XX transgenic animals. In the modified coding sequence, 306 nucleotide
XX positions have been replaced to lower the AT content (from 76 to
XX 49.7%) and to eliminate 10 mRNA instability motifs. The encoded
XX amino acid sequence is unaltered. In another modified sequence
XX (see AAK25593), a signal peptide sequence has been added and two
XX N-glycosylation sites eliminated. The invention provides modified
XX recombinant nucleic acid sequences and methods for increasing the
XX mRNA levels and protein expression of proteins that are difficult
XX to express in cell culture systems, mammalian cell culture systems
XX or in transgenic animals. The preferred difficult protein
XX candidates for expression are those derived from lower organisms
XX such as parasites, bacteria and viruses that have DNA coding
XX sequences of high AT content or which have mRNA instability motifs
XX or rare codons relative to the recombinant expression system to be
XX used. The invention allows expression of MSP-1 protein in the milk
XX of transgenic animals, and also provides a DNA vaccine comprising a
XX vector containing the altered MSP-1-42 sequence.

XX
SQ Sequence 1088 BP; 454 A; 139 C; 150 G; 345 T; 0 other;
Query Match 12.7%; Score 625.2; DB 20; Length 1088;
Best Local Similarity 74.3%; Pred. No. 3e-142;
Matches 789; Conservative 0; Mismatches 273; Indels 0; Gaps 0;
OY 3799 gctgtcactccttcgtgatgatgatacttcgtccaaatcgcgaagaatcgaagt 3858
Db 1 gcaagtaactcctccgttaattgatacaactcttcaaatgaaatgaaatcgaagt 60
OY 3859 cctatctgaaacccctcgcagcgctcatalagttccccaagaacagctggaatcac 3918
Db 61 ttattttaaacccttcgacgggtgttatatgaagttcaaaaaaacattgaaataac 120
OY 3919 gtgatgaccttcaatgttcaacgtggaagacattctgaaacggcgttcaataagagaa 3978
Db 121 gtaagacatttaagttaagttaagatattttaattccagattcaataaactgtaa 180
OY 3979 aattcaagaagcttttgagagcgacttgattcccttaagaagactgacctcctaac 4038
Db 181 aattcaaaaatggttttagaatcgaattcaattccataaagatttaacatcaagtat 240
OY 4039 taagtgtcaaggaaccatataagttccctcaataaagaagaaggaatctcgtct 4098
Db 241 tatgtgtcacaagaatccatataaatttcttaataaagaagaagaatctcctaagc 300
OY 4099 agttcaactatataagaagactccacgaacgatatacaatttcgtatgatgtgctg 4158
Db 301 agttaataatataagatccaatagatacgaataaatttgcgaatgatgttctt 360
OY 4159 gggatataagatctcgaagcaaaatacaagttcgccttgactgacttaataaagat 4218
Db 361 ggaataataaataatataccgaaataataaataagattagattcaataaataat 420
OY 4219 atcaacgataagcaagcgagaaatgaataatactgccttcctcgttaataacatcgaaac 4278
Db 421 atcaacgacaacaaagcgaaatgagaataatacctccctttaaacaataatgagacc 480
OY 4279 ctgtacaagaacgtgacgaacaaatcgacctcttcgttaattcacctgagagccaagtc 4338
Db 481 ttataataaacagttaaatgataaattgattattgttaattcaattgaagcaaaagt 540
OY 4339 ctcaacttacttaagagaagaagcaatgtggaagttaataacgaagctgaactc 4398
Db 541 ctaattatatacattagagaatcaacgtagaagttaaataaagaacttaattactta 600
OY 4399 aaaaacatccaagaacagctgacgatttcaagaagaaataaacaattcgtcgaattgca 4458
Db 601 aaacaattcaagaacaatttgcagattttaaaaaataacaatttcgttgaattgct 660
OY 4459 gacctgtctacgattataaccacaacatctcctgaccgaagttctgtccaatgcatg 4518
Db 661 gattatcaacagattataacataataactatgtgaacaagttccttagtaacagtatg 720
OY 4519 gtttgcgaaaactctgcgaaaacagtgctgcaattcgtcgcgaaggaactggaagcc 4578
Db 721 gttttgaaaactctgccaacacgctttatcttaattcattgtaggaacttgaaagt 780
OY 4579 atgcgaacatctcccaacgaacaatgctgtaagaagaacagttccccaatagcgcgt 4638
Db 781 atgttaaacatttccacaacacaaatgctgtaaaaaaacaatgtccacaatcttgcagt 840
OY 4639 ttcaagcatctgacgagcggaagagtgcgaagtgtctctcgtacatacaaaagaagga 4698
Db 841 ttcaagcatttagatgaagaagaagaaatgtaattgttttaattcaacaagaaggt 900
OY 4699 gataagtcgvgggaacccaacccctacgtcaatgtaaaaacatgggggtgtgagcc 4758
Db 901 gataaattgttgtaaaatccaatcctactctgttaacgaataatggtgtgagtca 960
OY 4759 gatgtaaatgcaccgaggaagacagcgctcctaacgaaagaataatcatatgcagagt 4818

Dh 961 gatcccaatgtaccgaagaattcagtagcaacgaaagaatacatgtaatgt 1020
Oy 4819 actaagccgactctatccactctctgacggatttttcg 4860
Dh 1021 actaaactgattcttaccacttcttgatgtattctgc 1062

RESULT 15
AAC68978

ID AAC68978 standard; DNA; 786 BP.

AC AAC68978;

DT 27-FEB-2001 (first entry)

DE Merozoite surface protein-133 coding sequence.

XX Merozoite surface protein; protazoacide; vaccine; malaria; ss.

OS Plasmodium falciparum.

PN W0200063245-A2.

PD 26-OCT-2000.

PE 20-APR-2000; 2000MO-GB013558.

PR 20-APR-1999; 99GB-0009072.

PR 13-MAY-1999; 99US-0311817.

PR 25-MAY-1999; 99CA-2271451.

XX (MEDT-) MEDICAL RES COUNCIL.

PI Holder A, Birdall B, Feeney J, Morgan W, Syed S, Uthairibull C;

DR WPI: 2001-015762/02.

XX P-PSDB; AAB37610.

PT Novel variants of the C-terminal fragment of Plasmodium merozoite

PS surface protein-1, useful as vaccines for treating or preventing

XX malaria.

XX Example 5; Fig 15; 126pp; English.

CC The present invention relates to non-natural variants of a C-terminal

CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The

CC non-natural variants have reduced affinity for at least 1 antibody

CC capable of blocking a second antibody that inhibits the proteolytic

CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least

CC one third antibody that inhibits the proteolytic cleavage of Plasmodium

CC MSP-1.4.2, compared to natural MSP-1.9. The non-natural variants of the

CC present invention are useful for immunising a mammal against malaria, and

CC can be used to treat malaria. The present sequence is the MSP-133

CC coding sequence.

XX Sequence 786 BP; 245 A; 164 C; 159 G; 218 T; 0 other;

Query Match 10.6%; Score 525.2; DB 22; Length 786;
Best Local Similarity 79.3%; Pred. No. 5.4e-118;
Matches 623; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

Dh 121 gtcatgacttcaacgcttaacgctcaagagacatttgaatccagattcaacaagaagaa 180
Oy 3979 aattcaagaagctcttgagagagacttgattcccttaagaaccttaacctctac 4038
Dh 181 aactcaagaagctcttgagagagacttgattcccttaagaaccttaacctctac 240
Oy 4039 tacgttgcaagaccctatacaagcttctcaataaagaagaagaaatttcgtc 4098
Dh 241 tacgttgtaagagaccctatacaagcttctcaataaagaagaagaaatttcgtc 300
Oy 4099 agttacaactatatacaagactccacacagatataatlttcgtatgatgtcgt 4158
Dh 301 tcttaactatcaatgaagactccatgatactgatactcaacttcgtcaagagctcgt 360
Oy 4159 gggtatcaagatctctgaggaagaataacagcttgccttgatctatataaagat 4218
Dh 361 ggttactcaagatctctgaggaagaataacagcttgccttgatctatataaagat 420
Oy 4219 atcaacgaagaagcgaagatgaataataatctgccttccgtgaataacatcgaac 4278
Dh 421 atcaatgataagcgaagagagaatgaataatgaatcctgcctgaataacatcgaac 480
Oy 4279 ctgtacaagaagctgaagacaataatcgacctcttgtaattcaacttgagagcaagtc 4338
Dh 481 ttgtacaagaagctgaagacaataatcgacctcttgtaattcaacttgagagcaagtc 540
Oy 4339 ctcaacttacttcaagagagagcaatgtggaagttaatacaagagctgaactactc 4398
Dh 541 ttgcatacacttcaagagagagcaatgtggaagttaatacaagagctgaactactc 600
Oy 4399 aaacaatcaagaagcttgcaagatttcaagaaatacaaatctgttggaattgca 4458
Dh 601 aagaccattcgaagagagcttgcaagatttcaagaaatacaaatctgttggaattgca 660
Oy 4459 gacctgtacagattatacaacacacacacacacacacacacacacacacacacac 4518
Dh 661 gatttgcacactgattatacaacacacacacacacacacacacacacacacacacac 720
Oy 4519 gtgttgaagaacctcgcgaagaacagctgagcaactctgctcgagcgaaccttgagc 4578
Dh 721 gtttcgagaaacttgctgaagagcttccctgcacacacacacacacacacacacac 780
Oy 4579 atgctg 4584
Dh 781 atgttg 786

Search completed: October 27, 2001, 15:53:08
Job time: 18524 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2001, 13:12:16 ; Search time 136.83 Seconds
(without alignments)
6834.736 Million cell updates/sec

Title: US-09-269-874-2

Perfect score: 4940

Sequence: 1 cgcacgcgatacgaatcatc.....ttcatcraatagatgatg 4940

Scoring table:

IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents: NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfill1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1265.2	25.6	5181	1 US-08-257-073-10	Sequence 10, Appl
2	74.2	1.5	7218	1 US-08-232-463-14	Sequence 14, Appl
3	60.6	1.2	5163	3 US-08-700-651-1	Sequence 1, Appl
4	60.6	1.2	5163	3 US-08-928-361B-4	Sequence 4, Appl
5	60.6	1.2	5318	3 US-08-700-651-2	Sequence 2, Appl
6	60.6	1.2	5318	3 US-08-928-361B-3	Sequence 3, Appl
7	58.8	1.2	5511	3 US-08-928-361B-2	Sequence 1, Appl
8	58.8	1.2	7334	3 US-08-928-361B-1	Sequence 2, Appl
9	49.6	1.0	1430	2 US-08-276-452A-25	Sequence 25, Appl
10	49.6	1.0	1430	2 US-08-798-744-25	Sequence 25, Appl
11	48.2	1.0	1690	1 US-08-276-452A-24	Sequence 24, Appl
12	48.2	1.0	1690	1 US-08-798-744-24	Sequence 24, Appl
13	46.4	0.9	198	5 PCT-US95-10668-3	Sequence 3, Appl
14	46.4	0.9	198	5 PCT-US95-10668-4	Sequence 4, Appl
15	45.8	0.9	340	5 US-08-182-175A-104	Sequence 104, App
16	45.8	0.9	340	5 PCT-US92-06412-104	Sequence 104, App
17	45.6	0.9	697	6 5171843-10	Patent No. 5171843
18	45.6	0.9	1137	6 PCT-US95-10668-1	Sequence 1, Appl
19	45.4	0.9	198	5 PCT-US95-10668-2	Sequence 2, Appl
20	45.4	0.9	198	5 PCT-US95-10668-3	Sequence 3, Appl
21	44.8	0.9	2223	1 US-08-257-073-4	Sequence 4, Appl
22	44.4	0.9	4766	5 PCT-US93-07261-10	Sequence 10, Appl
23	44.2	0.9	2277	1 US-08-676-967-2	Sequence 2, Appl
24	44.2	0.9	2277	1 US-08-676-974-2	Sequence 2, Appl
25	44.2	0.9	2277	1 US-09-098-487-2	Sequence 2, Appl
26	43	0.9	1241	1 US-08-471-033-39	Sequence 39, Appl
27	43	0.9	1241	1 US-08-471-033-42	Sequence 42, Appl

28	43	0.9	1241	2 US-08-471-044-39	Sequence 39, Appl
29	43	0.9	1241	2 US-08-471-044-42	Sequence 42, Appl
30	43	0.9	1241	2 US-08-463-483A-39	Sequence 39, Appl
31	43	0.9	1241	2 US-08-463-483A-42	Sequence 42, Appl
32	43	0.9	1241	2 US-08-471-046A-39	Sequence 39, Appl
33	43	0.9	1241	2 US-08-471-046A-42	Sequence 42, Appl
34	43	0.9	1241	2 US-08-470-566B-39	Sequence 39, Appl
35	43	0.9	1241	2 US-08-470-566B-42	Sequence 42, Appl
36	43	0.9	1241	2 US-08-469-334-39	Sequence 39, Appl
37	43	0.9	1241	2 US-08-469-334-42	Sequence 42, Appl
38	43	0.9	1241	3 US-09-300-529-39	Sequence 39, Appl
39	43	0.9	1241	3 US-09-300-529-42	Sequence 42, Appl
40	43	0.9	1358	1 US-08-471-033-45	Sequence 45, Appl
41	43	0.9	1358	2 US-08-471-044-45	Sequence 45, Appl
42	43	0.9	1358	2 US-08-463-483A-45	Sequence 45, Appl
43	43	0.9	1358	2 US-08-471-046A-45	Sequence 45, Appl
44	43	0.9	1358	2 US-08-470-566B-45	Sequence 45, Appl
45	43	0.9	1358	2 US-08-469-334-45	Sequence 45, Appl

ALIGNMENTS

RESULT 1
US-08-257-073-10
Sequence 10, Application US/08257073
Patent No. 5766597
GENERAL INFORMATION:
APPLICANT: Paoletti, Enzo
APPLICANT: de Taisne, Charles
APPLICANT: Tine, John A
TITLE OF INVENTION: MALARIA RECOMBINANT POXYVIRUS VACCINE
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Curtiss, Morris & Safford, P. C.
STREET: 530 Fifth Avenue, 25th Floor
CITY: New York
STATE: New York
COUNTRY: UNITED STATES OF AMERICA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,073
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/075,783
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/852,305
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,183
FILING DATE: 20-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2570
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066 CURTMS
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 5181 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-257-073-10

QY 2011 aagaagagatagacaacatgaagtgctcctgcccgaagtcgaagcctgatcaagaa 2070
Db 2098 AAAAAAGAGATGATTAATTAAGAAATTTATACCAAAAGTAAAGATGTTAA--GA 2155
QY 2071 gagaagaagaacatlaaaactgaagagacgtcagataaactcccgagccttcacagaaga 2130
Db 2156 AAGAACAGAGCTGCTTATTCAGATTTACACAACTTTAGTTGACAGCAACAACTCG 2215
QY 2131 gagaataccggaagcgtcaccacaagcccgagacaagcgcggttcagctcctgaagc 2190
Db 2216 AAGAT-GGGGGTCACTCCACACACATTAATCCAAATCAGAGAAACGAAAGTAAACAAA 2274
QY 2191 gatagcgtgcagctcgaagcgaagcagaagcagcagcgcctcagctgcagtgccc 2250
Db 2275 GAACAGAGAGAAAGAGAAAGAGTACGACACACAAACAGGTAAATACATTAACCA 2334
QY 2251 gttccagaagctaaagctcaagtgacctacacacacagctcctgtgaataaagaacgag 2310
Db 2335 CCAAAAGAGTAAAGTGTGAAATTCATTAAGAACATTAAGAGTAATGACAAATTCACAA 2394
QY 2311 aatgtcagcaacgtggaactcctgagaagctcctatgagctcccgataaactcctacatc 2370
Db 2395 GCCTTGACAAAGACAGTTATCTTAAAGAAATTAAGATGAATTTTAACTAAATCATATTA 2454
QY 2371 tgcacaataataatcctcgtcctcacaagcactatgaagcagaagatccttaaacagtac 2430
Db 2455 TGCTATTAATATTTATTTAGTATCAAACTCTAGTATGACCAAAATTTTATAGAGTATAT 2514
QY 2431 aagataaccaaagaagagagagtaaacgtcctcctgtgatacgaatgagcctgctgtc 2490
Db 2515 AATTTACTCCAGAGAGAAAGAAATGAATTAATTAATCATATGATTCATTTATTTATTT 2574
QY 2491 aatcgcagaacaacatcccgcttatcttatctctatgctcgtcagtagcctcaaatctctc 2550
Db 2575 AATTTTCAAAATTAACATACCTGCTATGATATCATTAATGATAGTAAACATGATTTA 2634
QY 2551 tctcaactgtcattgagatatacagagaagagatgctcgaactcgaactataaactcaaa 2610
Db 2635 CAACATCTCTTTTGAATTTATATCAAAAGAAATGATTTTATTTATCATTAACATAA 2694
QY 2611 gacaaagcaaatlaagaacctcctgaggaagcctaagaagctc----- 2657
Db 2695 GAGGAAATATCATCAAAAATTTATAGAGAGCAAAAACAAATTAACGAAATCATCATCT 2754
QY 2658 -----caacctgttaaacctcctc 2681
Db 2755 ACATTCAGTCTGGAAATACACCGTAAATCTGCTCAATCCGCAACTCACAGTAAATCC 2814
QY 2682 cagctcgaatgaacacactgtctcctcacacactcaagaaca----- 2720
Db 2815 CAANAACCAAAATCAATGCATCTCTACCAATACCCAATTTGGTGTAGTATCATCT 2874
QY 2721 -----gcccgaaagtgaagcgtcgaagcgaacactcctcgaacaaacttaataactca 2775
Db 2875 GGTCCTGCTGAGTGAAGAAAGTCATGATCCCTTAACAGATTTGCTATTAAGTAACGAT 2934
QY 2776 ctgaaactgtttggaacatcgtctcctcggaagaataaagaacat---ctaccaagaa 2832
Db 2935 TTGAAGAGTATTTGATGCTCTCTTAAATCTTGGAATAAATCAATTAAGTACCTTAATCCATTA 2994
QY 2833 ctatctgagaagaatcgtccgagaactctcagaagaatcactgaagaagcagagacaca 2892
Db 2995 ACCATTTCTACACAGAGATGGAATAATTTTATGAGATATTTTAAATAATTAATGATACC 3054
QY 2893 tctcataagagagctcgaactcgtggaatcctaaagcgaatgataaactcctc 2952
Db 3055 TATTTTATATGATATCAAAACATTCGTAAATCTAATTCAAAAAGTAAATTAACAGGTTTG 3114
QY 2953 aagaataatcctaagctaagaagcctggaagagacatcaataaagcgaagaagacactg 3012
Db 3115 ACC--GAAACACAAAAAAGCATTTAAATGATATAATTTAAAAAATTAAGATATCTTTA 3171
QY 3013 caactgagcttcgacgtctgatacaagaatacaactgaagaactgagagagactcctgacaag 3072

Db 3172 CAGTATTCATTTGATTTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAAG 3231
QY 3073 aagaagacagtcgagcaagataaagatcgaatcgaagaactgtgactgtcgaagagagag 3132
Db 3232 AAAAAAGAACTTGCGCAAGACAAAATGCAAAATTAATAAACTTACTTTATTAAGAACAA 3291
QY 3133 ctgaaagcaaacctaacctcgtcgaacaaatccgaacacgctacgtcgaactcctcagtg 3192
Db 3292 TTAGATCAAAATTTGAATTCCTTAATTAACCCACATTAATGATTAATTAATTAATTAAT 3351
QY 3193 tcttcaacaagaagaagccgagatcgcgagaagcagaagcaaacctcgtgagaaccc 3252
Db 3352 TTCTTTACAAAAAAGAAAGAACTTAATTAACAGAACTGAAACACATTTAGAAACACAA 3411
QY 3253 aagatcttccaacactacaagaagcctcgtlcaagatataaalygcagatcctcct 3312
Db 3412 AAAATATTATTGAAACATTAATAAGGACTTGTTAAATTTATTAATGATGAAATCATCTCA 3471
QY 3313 ctgaagactcctccgaggaagacatccagaaccgaggaactcgcagcctcgagaac 3372
Db 3472 TTAAAACTTTAAGTGAAGTATCAATTCAAACAGAGATTAATTAATTAATTAATTAATTA 3531
QY 3373 ttcaaggtcctgtcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 3432
Db 3532 TTTAGAGTATTAAGTAAATATGATGGAAGAACTCAATTAATTTATTAATTAATTAATTAAT 3591
QY 3433 aagctcagctcctcctcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 3492
Db 3592 AAATATCTTCTTATCAAGTGAGTATCATCATTTAATTAATTAATTAATTAATTAATTAAT 3651
QY 3493 aagaacaagaactcacccggaataagcccaagcgaagaaataaagaagtgataaagca 3552
Db 3652 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3711
QY 3553 ctggaatcttcaagaagaatctcgtcgtcgaagaagaatgctgcgaactgtgtgtc----- 3608
Db 3712 TTAATATCTTGAGAAATTTTCTCCACAGCAAAAGTTACAGAGTGTAACTCCACCT 3771
QY 3609 -----tgaatcgtgtcgaacacactgagcagctcga 3641
Db 3772 CAACGAGTGAATCCATCTCCATTAATCTGTAAGGGTAACTGAGTCAAGATCCACA 3831
QY 3642 aactaagaag-----cctgcactcactcaltgtcgaagcgaagcccaatataacc 3693
Db 3832 AAAGAGAAACAAATTAACCAATTCAGGCTCTTATTAACAGAAATTAACCAACAGTATGA 3891
QY 3694 acatctcagaagcgtcgaagatgaggtcgaatgagcgtcgaatgagcgtcgaatgagcgtcga 3753
Db 3892 CAATTTACAAATTTATGACGAGAGAGATGATCTTACGTTGATTAATTAATTTTGGAGAA 3951
QY 3754 agcgaaggaactacgaatgacactcgcgaagtggtcaccggtgagcgtcgaactcctcc 3813
Db 3952 TCCGAGAGTATGACGATATTTAGATCAATAGTAACTGGAGAAAGC--AATATCTGTC 4008
QY 3814 gtgattgataactcgttccaataacgaagaagaatacgaagcgtcctatcctgaact 3873
Db 4009 ACAAATGATTAATTCCTCTCGAGATTGAATAATGAATGATTAATTAATTAATTAATTA 4068
QY 3874 ctggaagcgtcctataggtcctcgaagaagcgtggaagaatacgtgatacctcgaat 3933
Db 4069 TTACGCTGAGTATTAAGAGCTTTAAAAAACAATTTGAAAAAACATTTTACATTTAAT 4128
QY 3934 gtcacagtgaaagacatctgaaagccgcttataaagaagaanaatlcgaagaagcgtc 3993
Db 4129 TTAATATTGACGATATCTTAAATTCACGCTTTAAGAACAAAAATATTCTTGATGATGA 4188
QY 3994 ttggaagcgaactgattcctataaagactcgtcctcctcctcctcctcgtgtgcaagac 4053
Db 4189 TTAGATCTGATTTAATGCAATTTAATCAATTAATCTCAAAATGAATACATTAATTAATTA 4248
QY 4054 ccataaagcttccataaagaagaagggataaatctcgtcgtcgaatcaactatc 4113

Db	4249	TCATTTAAATTAATTGAATTAGACACAAAAAACACACTTTTAAAAAGTTACAAATATATA	4308
Qy	4114	aaggacccacacgcacccgaatcaaatctgcgaatgatgctgaggatattacaagatc	4173
Db	4309	AAAGAATCAGTAAATGATATTTAAATTTGGCAGGAGGATTAAGTTATATGAAAG	4368
Qy	4174	ctgagcgaataatacaagctcgaccttgacctatataaaglatatcaacgataag---	4230
Db	4369	GTTTATACGAAATTAATAGATGATTTTGAATCATATTTTTAAAAAGTTATCAAAAGAAAG	4428
Qy	4231	-----caa	4233
Db	4429	GAGAGTTCCCATCATCACCACCAACAACCTCGCTCACCGACAAAAACAGACGACAA	4488
Qy	4234	ggcagagatgaataatactgcgccttcgcgaatacatgcaaacctctacaagcagtg	4293
Db	4489	AAGAAGGAAGTAAAGTCTCTTCATTTTAAACAACATTTGACACTTTATACATTACTTA	4548
Qy	4294	aagcacaatatcgacccctctcgtaattcacctgagagccaaagctcccaactactaac	4333
Db	4549	GTTTATTAATAATTTGCATTTACTTAATTAACCTTAAGGCAAAATTAACGATGTGTAATGT	4608
Qy	4354	gagagagcgaatgtagaagttaaatccaagagctgtaactactccaataacatccaagac	4413
Db	4609	GAAAAAATGAGACACACTGTTTAAATTAACCTTAAGTATTAAGCAATTTGATATAC	4668
Qy	4414	aagctggcagatcttcaagaaataatacaatctgcgcgaattgcagacctgctacgat	4473
Db	4669	AAATTAATCTTTTAAAAACCATTAACGACTTGCAGACATTTAAATAATGATTAATGAT	4728
Qy	4474	tataacacacaacatctcctcgaccaaattctgccaactgacatgagtggttcgnaaacctc	4533
Db	4729	GATACGAAAAAAGATATGCTTTGGCAAAATTACTTAGTACAGGATTAG---TTCAAAATTTT	4785
Qy	4534	gccaaaacagtgctcgagcaatctgcctcgacaggaacctgcaggcatctgtaaatctcc	4553
Db	4786	CCTATATCAATAATATCAAAATTTAATTGAAGGAAATTTCCAAAGATATGTATTAACCTTCA	4845
Qy	4594	cagcacaatgctgtagaagaacagctgccccagaatagcgcctgtcttcaggatctgac	4653
Db	4846	CAACACCAATGCGTAAAAAACAATGTCAGAAATTTCTGCAATGTTTCAGACATTTAGAT	4905
Qy	4654	gagcgcgaagatgtagagtgctcctcgaaactaanaacaaagaaagagataaagtgcgtgag	4713
Db	4906	GAAAGAAAGAAAGTAAATGTTTATTAAATTTCAACAAGAGAGTGATAAATGCTTTGAA	4965
Qy	4714	aaccacaacctactcgtcaatgtaaacaaatgycgggtgtgtagccgcatgctcaatagcacc	4773
Db	4966	AATCCAATTCCTACTTGTATAACGAAATTAATATGGTGATGTAAGTACAGATGCCAAATGATGC	5025
Qy	4774	gaggaaagacagcgcctctaaacggaaagaataacaaatgtagatgtgactaaagccgaccc	4833
Db	5026	GAAAGAGTTTCAGTACCAACGGAAGAAATACATGTGGAATGTGACTTAACCTGATTTCT	5085
Qy	4834	tatccactctcgacgggaatttttgcctcagcttaatttcttggagcatctccctcgt	4893
Db	5086	TATCCACTTTTCGATGCTATTTTCTCGAGTCCCTCTTAACCTTCTTGAGATATCATTTTAA	5145
Qy	4894	ctgatctcatgctgcatcccgtaacagctcatctaa	4929
Db	5146	TTAATACCATGTTAAATTAATTAATACGTTTCAATTA	5181

RESULT 2
 US-08-232-463-14/c
 Sequence 14, Application US/08232463
 Patent No. 5670367
 GENERAL INFORMATION:
 APPLICANT: DORNER, F.
 APPLICANT: SCHEIFELINGER, F.
 APPLICANT: FALKNER, F. G.
 TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
 NUMBER OF SEQUENCES: 52

	CORRESPONDENCE ADDRESS:	
	ADDRESSEE: Foley & Lardner	
	STREET: 1800 Diagonal Road, Suite 500	
	CITY: Alexandria	
	STATE: VA	
	COUNTRY: USA	
	ZIP: 22113-0299	
	COMPUTER READABLE FORM:	
	MEDIUM TYPE: Floppy disk	
	COMPUTER: IBM PC compatible	
	OPERATING SYSTEM: PC-DOS/MS-DOS	
	SOFTWARE: PatentIn Release #1.0, Version #1.25	
	CURRENT APPLICATION DATA:	
	APPLICATION NUMBER: US/08/232,463	
	FILING DATE:	
	CLASSIFICATION: 435	
	PRIOR APPLICATION DATA:	
	APPLICATION NUMBER: US/07/935,313	
	FILING DATE:	
	APPLICATION NUMBER: EP 91 114 300.6	
	FILING DATE: 26-AUG-1991	
	ATTORNEY/AGENT INFORMATION:	
	NAME: BENT, Stephen A.	
	REGISTRATION NUMBER: 29,768	
	REFERENCE/DOCKET NUMBER: 30472/114 IMMU	
	TELECOMMUNICATION INFORMATION:	
	TELEPHONE: (703)836-9300	
	TELEFAX: (703)683-4109	
	TELEX: 899149	
	INFORMATION FOR SEQ ID NO: 14:	
	SEQUENCE CHARACTERISTICS:	
	LENGTH: 7218 base pairs	
	TYPE: nucleic acid	
	STRANDEDNESS: single	
	TOPOLOGY: linear	
	IMMEDIATE SOURCE:	
	: CLONE: PTZgpt-F1s	
	US-08-232-463-14	
Query Match	1.5%; Score 74.2;	DB 1; Length 7218;
Best Local Similarity	7.8%; Pred. No. 2e-09;	
Matches	34; Conservative	234; Mismatches 167; Indels 0; Gaps 0;
OY	1693 aaagaactaatatattacaagaatctcaaaagttaagtcgaacgcgtt	1752
	: : : : : : : : :	
Dd	1465 AAGTAGTTAAAGAATGAGAGATTTCGTACRRRRRRRRRRRRRRRRR	1406
OY	1753 gttggaacattaagaagatagaagaacgttgttgtagaagaagatac	1812
	: : : : : : : : : : : : : : : : : : :	
Dd	1405 RRR	1346
OY	1813 aataaccagatgagaaagtcctgtgagtcctcgatatgttaaagtc	1872
	: : : : : : : : : : : : : : : : : : :	
Dd	1345 RR	1286
OY	1873 gtgtccctcatbaacaagatgatgaacctcaaagaacccaactatct	1932
	: : : : : : : : : : : : : : : : : : :	
Dd	1285 RR	1226
OY	1933 gagttaaaaataatacatatgltgccgaalagtlaaagcaggagataag	1992
	: : : : : : : : : : : : : : : : : : :	
Dd	1225 RR	1166
OY	1993 tactactcatcgtlactcaagaagaagataagcataaagtgttcalgc	2052
	: : : : : : : : : : : : : : : : : : :	
Dd	1165 RR	1106
OY	2053 gaggagcctgatcaagaagaagaagaacatlaaacgaagagacagat	2112
	: : : : : : : : : : : : : : : : : : :	
Dd	1105 RRTGC	1046
	AAGCTCCCGACTGG 1046	
OY	2113 gagctctcacagaa 2127	

OY		gcaaggcgagatgaanaaatctgcgttcgccattcagaataacatcgaaaccctggtaaacg	4289
Db	914	AGAAACCAACAACAACACTGATCACAACAACAACAACAACAACAACAACAACAACA	973
OY	4290	agttaacgacaanaatcgacctcttctglaatlcaactcgtagggccaagtgccctaactlac	4349
Db	974	CCACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	1033
OY	4350	ttaaagaaaagagcatabtgtgaagttaatcaaggagtcgaaetcatcccataaacatlc	4409
Db	1034	CACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	1093
OY	4410	agacacatcgtagttcacagaanaatatcaaatlctgcggaattlgaaactgttac	4469
Db	1094	CAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	1153
OY	4470	cgatataaccaacaacatltcc	4492
Db	1154	CAACTACTTACTACTCTAACGC	1176

```

RESULT 5
US-08-700-651-2
; Sequence 2, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5318
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-08-700-651-2

```

Query Match	1.2%;	Score 60.6;	DB 3;	Length 5318;
Best Local Similarity	44.2%;	Pred. No. 6.7e-06;		
Matches 249;	Conservative	0;	Mismatches 314;	Indels 0;
			Gaps	0;

OY	3930	caatgtaacgtgaagacgtcttgcagcgcgctttaatgaagaaatttcaaga	3989
Db	615	caacaacaacaacaagaacaacaacactactactaaccctactattacga	674
OY	3990	cgctctggagcgacttga ttccctataagaagccgcgcctctctaaactgtgtcaa	4049
Db	675	caacaacaacaacaacaacaacaacaacacacacacacacacacacacacaa	734
OY	4050	ggaccatacaagtctctcaataaagaagaggataattctgtctagttaacata	4109
Db	735	cgactactactactactactactactactactactactactactaacaacaacaacta	794
OY	4110	tataaaggatccatgcagacgcatatcaatttcgttaatgtagtgcggygtattcaa	4169
Db	795	caaccacaactacacaacaacaacactacacccaacaacaacaacacaactataa	854
OY	4170	gatctttagcgaaaaatatacaagtctgcacttgccttatataaaagtlataaagataa	4229
Db	855	ccacaactatcacacaacacacactacacccaacaacacacacacacccaactaccca	914
OY	4230	gcaagcgagaaatgaaaataatctgccttctctgataacatcgaaacccgttacaagac	4289

Db	915	agaaacacacacacacacactctacaacacacacacacacacacacacactactactataacaca	974
Oy	4290	agltgaagacaanaatcgaccttctgtaalttcactgtgagcgcaagttccttaactatac	4399
Db	975	ccacacacacacacacacacacacactcaactcactcagaacacacacactactactacta	1034
Oy	4350	ttacgagagagacatgtggaagtttaaatcagagagtcgaactactcctaaacatcca	4409
Db	1035	ctaccacacacacacacacactactactactcctccacacacacacacactactactacta	1094
Oy	4410	agacaaagctgycagatttcaagaaaataacaaatttcgtcgaaatlycagacctgtctac	4468
Db	1095	caacacacacacacacacacacacacacacacacacactcaccgaaaccaaacacacacaca	1154
Oy	4470	cgattatnaccacacacatctcc	4492
Db	1155	caactactactactactactacaac	1177

```

: RESULT 6
: US-08-928-361B-3
: Sequence 3, Application US/08928361B
: Patent No. 6071518
: GENERAL INFORMATION:
: APPLICANT: Petersen, Carolyn
: TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
: TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
: TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
: TITLE OF INVENTION: SPECIES INFECTIONS
: NUMBER OF SEQUENCES: 30
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PETERS, VERNY, JONES & BIKSA
: STREET: 385 Sherman Avenue, Suite 6
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94306-1840
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/928,361B
: FILING DATE: 12-SEP-1997
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/026,062
: FILING DATE: 13-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Verny, Hana
: REGISTRATION NUMBER: 30,518
: REFERENCE/DOCKET NUMBER: 480.76-1(HV)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-324-1677
: TELEFAX: 650-324-1678
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5318 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-928-361B-3

```

	Query Match	1.28;	Score 60.6;	DB 3;	Length 5318;
	Best Local Similarity	44.28;	Pred. No. 6.7e-06;		
	Matches 249;	Conservative	0;	Mismatches 314;	Indels 0;
					Gaps
OY	3930	caatgtcaagctgaagacattctgaacagcgctttaatagaagaaattcaagaa			3988
Db	614	CACACCAACACACAGCACAACAACAACACTACTACTACTACTACTACTACTACGA			673

QY 3990 cgtcttgagagcgactgtatccctataaagacgtgacctctcttaactacgtgtca 4049
Db 674 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 733
QY 4050 ggaccatacaagttctcctaataagagaagaggaattctgtctagtaacta 4109
Db 734 CGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 793
QY 4110 tatcaaggacccatcgcaccgcgatataatctgcctaagtgtgcgggtataca 4169
Db 794 CAACCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 853
QY 4170 gatcctgagcgaataatacaagctcgtacctgtactataaagaatatacgaata 4229
Db 854 CCACAAGTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 913
QY 4230 gcaagcgagagatcaaaataatctgacctctcctgaataacatcgaaacctgtacaagac 4289
Db 914 AGAAGCCAAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 973
QY 4290 agtgaagcgaataatgacctctcgtatcactgagcgaagtcctcaactatac 4349
Db 974 CCACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1033
QY 4350 ttacgagagagcgaatgtggaagttaaatcaagagctgactacccaacaacatca 4409
Db 1034 CTACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1093
QY 4410 agacaagctgcagattcgaagaataatacaattcgtcggaattcgagacctgtctac 4469
Db 1094 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1153
QY 4470 cgattataaccaacaacatctcc 4492
Db 1154 CACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 1176

RESULT 7
US-08-928-361B-2
Sequence 2, Application US/08928361B
Patent No. 6071518
GENERAL INFORMATION:
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESSES:
ADDRESSEE: PETERS, VERNY, JONES & BIKSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: VERNY, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480,76-1(HV)
TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5511 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-928-361B-2

Query Match 1.2%; Score 58.8; DB 3; Length 5511;
Best Local Similarity 45.2%; Pred. No. 2,1e-05;
Matches 216; Conservative 0; Mismatches 267; Indels 0; Gaps 0;

QY 3936 caacgtgagagacattctgaacgcgcgtttaataagagagaaatttcaagaagctt 3995
Db 1094 CCAGCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 1153
QY 3996 ggaagcgactgtatccctataaagacctgacctccttaactacgtgtcgaagacc 4055
Db 1154 CACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 1213
QY 4056 atacaagttcctcaataaagaagaggaataattctgtctagttacaactataca 4115
Db 1214 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 1273
QY 4116 ggaactcagacacgcgatataatctgctaataatgtgtgtggtatcaagaatcct 4175
Db 1274 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1333
QY 4176 gagcgaataatacaagctgcacctgtactctataaagaatatacaagtaagcagg 4235
Db 1334 CAACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1393
QY 4236 cgaagatgaataatctgcacctctcctgaataacatcgaaacctgtacaagagtgaa 4295
Db 1394 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1453
QY 4296 cgacaataatcgacctctcgttaattcacctgagagcgaagtcctcaactaactaaga 4355
Db 1454 CCACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1513
QY 4356 gaagagcagatgtgaagttaaatcaagagctgaactactcaaaacatccaagac 4413
Db 1514 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 1571

RESULT 8
US-08-928-361B-1
Sequence 1, Application US/08928361B
Patent No. 6071518
GENERAL INFORMATION:
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESSES:
ADDRESSEE: PETERS, VERNY, JONES & BIKSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```

1  APPLICATION NUMBER:  US/08/928,361B
2  FILING DATE:  12-SEP-1997
3  CLASSIFICATION:
4  PRIOR APPLICATION DATA:
5  APPLICATION NUMBER:  US 60/026,062
6  FILING DATE:  13-SEP-1996
7  ATTORNEY/AGENT INFORMATION:
8  NAME:  VERNY, Hana
9  REGISTRATION NUMBER:  30,518
10 REFERENCE/DOCKET NUMBER:  480.76-1(HV)
11 TELECOMMUNICATION INFORMATION:
12 TELEPHONE:  650-324-1677
13 TELEFAX:  650-324-1678
14 INFORMATION FOR SEQ ID NO:  1:
15     SEQUENCE CHARACTERISTICS:
16     LENGTH:  7334 base pairs
17     TYPE:  nucleic acid
18     STRANDEDNESS:  double
19     TOPOLOGY:  linear
20     MOLECULE TYPE:  DNA (genomic)
21     OS-08-928-361B-1

```

Query Match	1.2%	Score 58.8	DB 3	Length 7334
Best Local Similarity	45.2%	Pred. No. 2.4e-05		
Best Match 216, Conservative	0	Mismatches 262	Indels 0	Gaps 0

Qy	3936	caagtgaaagacatctctgaacgcgcgctttaatgaagagaaattccaagaagctct	3995
Dp	2762	CCAGGCAACAACCTACAAACCAACCAACAACTACCAAGAAACCAACAAACAAACAA	2821
Qy	3996	gagagcgacttgattccctataaagacctgacctctcttaactacgtgttcaagacc	4055
Dp	2822	CAACTACTACTACTACAAACAACAACAACAACACTACTACTACTACTACTACTACTA	2881
Qy	4056	atacaagttcctcaataaagaagaaggataattctgctagttaacaataataa	4115
Dp	2882	CTACTACTACTACTAGTACTACCAACAACAACAACAACAACAACAACAACAACAACA	2941
Qy	4116	gactcatcgacacgcgatalcaatttcgtaatagtgtcgtgggattacaagctct	4175
Dp	2942	CACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	3001
Qy	4176	gagcgaaaaatacaagtcgcaccttgactctatlaaaagtalcaagataagcaag	4235
Dp	3002	CAACCAACAACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	3061
Qy	4236	cgagaatgaaaataatctgcacctcttgataatacatcgaaacctgttcaagacgtga	4295
Dp	3062	CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	3121
Qy	4296	cgacaaaatcgacctctctglaataccttggagcgacaggtctccaactaactacga	4355
Dp	3122	CCACAACAACAACCGCAACCAACAACACTACCAAGAAACCAACAACAACAACAACA	3181
Qy	4356	gaagagccaatgltgaagttlaaatcaagagcttgaactacctccaacaactcaagac	4413
Dp	3182	CTACTACTACTACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	3239

RESULT 9
US-08-276-452A-25
Sequence 25, Application US/08276452A
Patent No. 5646029
GENERAL INFORMATION:
APPLICANT: Chen, Chao-Guang
APPLICANT: Mau, Shiao-Lim
APPLICANT: Du, He
APPLICANT: Gane, Alison M
APPLICANT: Bacic, Antony
APPLICANT: Clarke, Adrienne E
TITLE OF INVENTION: Plant Arabinoxylactan Protein (Acp) Genes
NUMBER OF SEQUENCES: 91

CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: United States of America
ZIP: 80503

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,452A
FILING DATE: 18-JUL-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Caruthers, Jennie M.
REGISTRATION NUMBER: 34,464
REFERENCE/DOCKET NUMBER: 27-91A

TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)499-8080
TELEFAX: (303)499-8089
TELEX: 496617824

INFORMATION FOR SEQ ID NO.: 25:
SEQUENCE CHARACTERISTICS:

```

? NAME/KEY: CDS
? LOCATION: 2..1312
?
? FEATURE:
?
? NAME/KEY: misc_feature
? LOCATION: 41..112
?
? OTHER INFORMATION: /note="Derived amino acid sequence
? corresponding to the peptide sequence by protein
? OTHER INFORMATION: microsequencing"
?
? FEATURE:
?
? NAME/KEY: misc_feature
? LOCATION: 41..112
?
? OTHER INFORMATION: /note="Derived amino acids 14-24,
? 28, 30, 32-37 are identical to sequences obtained
? OTHER INFORMATION: by protein microsequencing"
?
? FEATURE:
?
? NAME/KEY: misc_feature
? LOCATION: 25..31
?
? OTHER INFORMATION: /note="Amino acids 25, 27, 29, and
? OTHER INFORMATION: 31 are hydroxylated proline residues; amino acid
? OTHER INFORMATION: 26 can be T instead of A"
?
? OS-08-276-452A-25

```

Query Match	1.0%	Score 49.6;	DB 1;	Length 1430;
Best Local Similarity	44.3%;	Pred.	0.0025;	
Matches 202;	Conservative	0;	Mismatches 254;	Indels 0;
				Gaps 0;

[illegible]


```

Db 701 AACCAACAACGGCTTCTCCGAGAGTTTACAACAGACAACAACAACAATFTTTTC 760
Qy 3325 tcgaggaagagcatccagaccgagagtaactacgccagcttcgagaactcaaggctcgy 3384
Db 761 TCCGCAATTAACAACAACATAATAATTAACAATAATGTTTCTCCGAGATTACAACAAT 820
Qy 3385 tctgaagtcgaaggaagcgaagctggaagacaactcgaaccttggaagaagaagtcgcctac 3444
Db 821 AATAACAATAATGTTTCTTCGCCGAAATTAACAACAAATTAACAACAATGCTTTCGCCGAG 880
Qy 3445 ctcctagcgcagctgcacccaacctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3504
Db 881 AACTACAACAACAATATATGTTTCTTCGCCGAAATTAACAACAAAACAATTAACAACAATGCT 940
Qy 3505 tacacgcgcgaatgcccgaagcgcgaataataacagac 3540
Db 941 TTCTGTGAGATTACAACAACAACAACAACAATGCC 976

RESULT 10
US-08-798-744-25
: Sequence 25, Application US/08798744
: Patent No. 5830747
:
: GENERAL INFORMATION:
: APPLICANT: Chen, Chao-Guang
: APPLICANT: Mau, Shiao-Lim
: APPLICANT: Du, He
: APPLICANT: Gane, Alison M
: APPLICANT: Bacic, Antony
: APPLICANT: Clarke, Adrienne E
: TITLE OF INVENTION: Plant Arabidogalactan Protein (AGP) Genes
: NUMBER OF SEQUENCES: 91
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Greenlee and Winner, P.C.
: STREET: 5370 Manhattan Circle, Suite 201
: CITY: Boulder
: STATE: Colorado
: COUNTRY: United States of America
: ZIP: 80303
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/798,744
: FILING DATE: 13-FEB-1997
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/276,452
: FILING DATE: 18-JUL-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Caruthers, Jennie M.
: REGISTRATION NUMBER: 34,464
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (303)499-8080
: TELEFAX: (303)499-8089
: TELEX: 49617824
: INFORMATION FOR SEQ. ID NO.: 25:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1430 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2..1312
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: 41..112
: OTHER INFORMATION:
: /note= "Derived amino acid sequence

```

```

: OTHER INFORMATION: corresponding to the peptide sequence by protein
: OTHER INFORMATION: microsequencing"
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: 41..112
: OTHER INFORMATION: /note= "Derived amino acids 14-24,
: OTHER INFORMATION: 28, 30, 32-37 are identical to sequences obtained
: OTHER INFORMATION: by protein microsequencing"
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: 25..31
: OTHER INFORMATION: /note= "Amino acids 25, 27, 29, and
: OTHER INFORMATION: 31 are hydroxylated proline residues; amino acid
: OTHER INFORMATION: 26 can be T instead of A"
: US-08-798-744-25

Query Match 1.0%; Score 49.6; DB 2; Length 1430;
Best Local Similarity 44.3%; Pred. No.0.0025;
Matches 202; Conservative 0; Mismatches 254; Indels 0; Gaps 0

QY 3085 ggcaagataaagatgcagatacaagaagttgactctgctcaagagcagcttgaagcaaa 3144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 521 GGGTACTCGGAGAAATATACACACACAAAGAACAAATGGTACTCGAGAAATATACAAAC 580
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3145 ctcaactcactgaacaalccgaacacgctactgcgaactctcagtygtcttcaacaag 3204
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
DB 581 AACACACAAATAGGCTACTCCGAGAAATATACACACAAACAAATAGGCTACTCCAAAGAT 640
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 3205 aagaaggaagccgagatgcgcgagacagaagaacactctggagagaacccaagttcttc 3264
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
DB 641 TGCACACAAATAGGCTACTCCAAAAAATCAACAAATATAGCTTCTCCGAAATATACATG 700
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 3265 aaacctacaagagcctctgtaagattatataatggcgagctcttcctcttgaagactctc 3324
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
DB 701 AACACACAAACAGGCTCTCCGAGAGTTTCAACACAGCACACACAAACAAACAAATTTTTC 760
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 3325 tccggaggaagactcagacccgagagataactacgcgcagctcgaagaacttcaaggtcc 3384
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
DB 761 TCCGAGAAATATACACAAATATATCAATATATGTTTCTCCGAGAAATATACAAACAAAT 820
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 3385 tctaaagtcgaaggaagctgaagacaaactgaaactggagagaagaagactcagctac 3444
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
DB 821 AATACAAATATATGTTTCTCCGAGAAATATCAACAAACAAATATACAAACAAATGCTTCCGAG 880
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 3445 cctctagcgagctgcatcactgctgcgcgagctcgaagaagctcattaagaagaagaac 3504
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
DB 881 AACTACACAAACAAATATATGTTTCTCCGAGAAATATCAACAAACAAACAAATATACAAATGCT 940
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 3505 tacaccgcaatagcccaagcgaagaataataacagac 3540
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
DB 941 TTCTCTGAGAAATATACAAACAAACAAACAAATATGCTC 976
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 11
US-08-276-452A-24
: Sequence 24, Application US/08276452A
: Patent No. 5646029
: GENERAL INFORMATION:
: APPLICANT: Chen, Chao-Guang
: APPLICANT: Mau, Shao-Lim
: APPLICANT: Du, He
: APPLICANT: Gane, Alison M
: APPLICANT: Bacic, Antony
: APPLICANT: Clarke, Adrienne E
: TITLE OF INVENTION: Plant Arabidnagalactan Protein (AGP) Genes
: NUMBER OF SEQUENCES: 91
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Greenlee and Winer, P.C.
: STREET: 5370 Manhattan Circle, suite 201
: CITY: Boulder
: STATE: Colorado
: COUNTRY: United States of America

```

```

ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276.452A
FILING DATE: 18-JUL-1994
CLASSIFICATION: A35
ATTORNEY/AGENT INFORMATION:
NAME: Caruthers, Jennie M.
REGISTRATION NUMBER: 34,464
REFERENCE/DOCKET NUMBER: 27-91A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)499-8080
TELEFAX: (303)499-8089
TELEX: 49617824
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1690 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 60..1442
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..38
OTHER INFORMATION: /note= "Nucleotide sequence
OTHER INFORMATION: obtained by PCR which does not overlap with the
OTHER INFORMATION: cDNA clone"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 60..128
OTHER INFORMATION: /note= "Predicted, transmembrane
OTHER INFORMATION: segment"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 135..179
OTHER INFORMATION: /note= "Derived amino acid sequence
OTHER INFORMATION: corresponding to the peptide sequence by protein
OTHER INFORMATION: microsequencing"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 135..179
OTHER INFORMATION: /note= "Amino acids 27 to 36, 38,
OTHER INFORMATION: and 40 are identical to that in the peptide
OTHER INFORMATION: obtained by direct microsequencing"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 135..179
OTHER INFORMATION: /note= "Amino acid 26 may also be
OTHER INFORMATION: determined by direct microsequencing"
OTHER INFORMATION: Ala; 37 and 39 can also be determined residues"

```

[illegible]

Db	693	GGCTACTCGGAGAAATGCTAATAATACAAAAATTAACAAATGCTACTCAGAGAAATTAACAACAC	752
Qy	1039	aacaagaagatagagagagccacgagaagagatlcuaagagatcggcaaaacatlaagtc	1098
Db	753	AATTAACAACATGGCTTACGCCAGAAATTAACACAAATGGCTCTCAGAGTTACACAAC	812
Qy	1099	aacatagattctctcttacttgaatcccttgagctggtactacttgaagagaagat	1158
Db	813	AACATATATTTTACTCTCGGGAATTAACAACACACACACAAATGTTTCTCGGAGAT	872
Qy	1159	aagaatatagacatctccgcgaagtcgagacaagaatlcacacgaacctaagat	1218
Db	873	TTCACAAACAATGGCTTACTCCAAAAGATCAACAAATATGGCTACTCCGAGATTATACATG	932
Qy	1219	cccaagtgatgacgtacccctgtgcttataacgatalcaacaacgctctcaagagctc	1278
Db	933	AACACACACATGGCTTCTCGAGAGTTTCAACACACACACACACACACACACAC	992
Qy	1279	aatagcttggtgactctgattaaaccccttcgattatagcaagaacccctcaagatatc	1338
Db	993	AACGTTTTCTCTGGAATTAACAACAACAACAATAAATAAATGTTTCTCCGAGATTATC	1052
Qy	1339	tacacgacatgagagaagaagtatlcacagaacatcaagagaagatca	1391
Db	1053	AACACACACATATACACAAATGCTTCTCTCAGAGAAATTAACAACAACACACAA	1105

RESULT 12
 US-08-798-744-24
 Sequence 24, Application US/08798744
 Patent No.5830747
 GENERAL INFORMATION:
 APPLICANT: Chen, Chao-Guang
 APPLICANT: Mau, Shiao-Lim
 APPLICANT: Du, He
 APPLICANT: Gane, Alison M
 APPLICANT: Bacic, Antony
 APPLICANT: Clarke, Adrienne E
 TITLE OF INVENTION: Plant Arabidogalactan Protein (AGP) Genes
 NUMBER OF SEQUENCES: 91
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Greenlee and Winner, P.C.
 STREET: 5370 Manhattan Circle, Suite 201
 CITY: Boulder
 STATE: Colorado
 COUNTRY: United States of America
 ZIP: 80303
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/798,744
 FILING DATE: 13-FEB-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/276,452
 FILING DATE: 18-JUL-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Caruthers, Jennie M.
 REGISTRATION NUMBER: 34,464
 REFERENCE/DOCKET NUMBER: 27-91A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (303)499-8080
 TELEFAX: (303)499-8089
 TELEX: 49617824
 INFORMATION FOR SEQ ID NO: 24:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1690 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

APPLICANT: James Eberwine
TITLE OF INVENTION: A Method of Sequencing Proteins by
TITLE OF INVENTION: Epitope Ordering and Protein
TITLE OF INVENTION: Restriction Mapping
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jane Massey Licata, Esq.
STREET: 210 Lake Drive East, Suite 201
CITY: Cherry Hill
STATE: NJ
COUNTRY: USA
ZIP: 08002
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM 486
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10668
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/294,133
FILING DATE: August 22, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jane Massey Licata
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: PENN-0137
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 779-2400
TELEFAX: (609) 779-8488
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 198
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: No
PCT-US95-10668-4

Query Match 0.9%; Score 46.4; DB 5; Length 198;
Best Local Similarity 52.6%; Pred. No. 0.0054;
Matches 101; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
QY 1357 aagaatttcaacgaatcaagagagatccaattggaagaagaattgaggt 1416
DB 2 AACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 61
QY 1417 gacagaagaattacgaagacgcagcaaaagtctaacgatatacctaagaagtga 1476
DB 62 AACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 121
QY 1477 aagctgtgaacgagatctatgatctcaaatcaacaataacatgacctgacaaattc 1536
DB 122 AACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 181
QY 1537 gagaataatgatg 1548
DB 182 AAGCTGATGCTG 193

RESULT 15
US-08-182-175A-104
Sequence 104, Application US/08182175A
Patent No. 5559223
GENERAL INFORMATION:
APPLICANT: Saverio Carl Falco
APPLICANT: Sharon J. Keeler
APPLICANT: Janet A. Rice
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing F
NUMBER OF SEQUENCES: 113
CORRESPONDENCE ADDRESS:

ADDRESSEE: E.I. du Pont de Nemours and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/182,175A
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/743,006
FILING DATE: 9 August 1991
ATTORNEY/AGENT INFORMATION:
NAME: Linda Axamethy Floyd
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-4929
TELEFAX: (302) 892-7949
TELEX: 835420
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 340 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
STRAIN: E. coli
CELL TYPE: DHS alpha
IMMEDIATE SOURCE:
CLONE: segment 534 [seg 534]
FEATURE:
NAME/KEY: CDS
LOCATION: 3..326
OTHER INFORMATION: /function= "synthetic seed storage protein"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "ssp"
OTHER INFORMATION: /standard_name= "ssp-534"
US-08-182-175A-104

Query Match 0.9%; Score 45.8; DB 1; Length 340;
Best Local Similarity 49.0%; Pred. No. 0.011;
Matches 122; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
QY 2973 gaagctggaagagacatcaataagctgaagaagacactgcaactgagcttcgactgta 3032
DB 62 GAAGCTGAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAT 121
QY 3033 caacagtaacaactgaacactggaagagactcttcgacaagaagaagacagtcggaagta 3092
DB 122 GAAGAAAGCTGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 181
QY 3093 taagatcagatcaagaagctgactctgctcaagagagacgttgaagaacaactcaactc 3152
DB 182 GATGAAGAAGCTCGAAGAGAGAGATGAAGCTCATGAGAGAGAGAAAGATGAAGAAAGCTGAAGA 241
QY 3153 actgacaatccgaagaacgactgcaagaacttcacgtgtcttcacaagaagaaga 3212
DB 242 AAGATGAAGCAATGAAGCAAAATGAAGTGGCTTGAGAGAGAAAGAAAGAGCTGA 301
QY 3213 agccgagat 3221
DB 302 AGAGAGAT 310

Mon Oct 29 07:52:14 2001

us-09-269-874-2.rni

Page 13

Search completed: October 27, 2001, 13:13:25
Job time: 14001 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2001, 13:10:47 ; Search time 8587.58 Seconds
(without alignments)
8897.813 Million cell updates/sec

Title: US-09-269-874-2
Perfect score: 4940

Sequence: 1 ccgacgcgcatgataaatcat.....ttcatctaatagatgatgg 4940

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4940	100.0	4940	9 A92451	A92451 Sequence 2
2	4938.4	100.0	4940	56 CPE131294	A1131294 Cloning v
3	2816.4	57.0	5917	96 PFANT195	X02919 Plasmodium
4	2816.4	57.0	7038	96 PFEP195A	X15063 Plasmodium
5	2814.4	57.0	4920	9 A92450	A92450 Sequence 1
6	2706.2	54.8	5760	9 A04562	A04562 P.falciparu
7	2706.2	54.8	5760	10 E00656	E00656 cdna encodi
8	2686.8	54.4	5282	96 PFP190	X03371 P.falciparu

9	1400.6	28.4	5271	5	AF062349	AF062349 Plasmodium
10	1331.2	26.9	5243	5	AF062348	AF062348 Plasmodium
11	1330.8	26.9	5085	5	AF218248	AF218248 Plasmodium
12	1290.4	26.1	5392	96	PF0190G1	PF0190G1 Plasmodium
13	1290.4	26.1	5754	96	PFPMMSA1	M9143 P. falcipar
14	1272.2	25.8	5312	96	PFMEZ5A1B	M35327 P. falcipar
15	1267.2	25.7	5276	96	PFPGP195	M37213 P. falcipar
16	1267.2	25.7	5541	96	PFPGP195	M30831 P. falcipar
17	910.4	18.4	1636	96	PFPMSPB06	D13357 P. falcipar
18	910.4	18.4	1636	96	PFPMSPB08	D13358 P. falcipar
19	908.8	18.4	1636	96	PFPMSPB37	D13360 P. falcipar
20	908.8	18.4	1636	96	PFPMSPB37	D13362 P. falcipar
21	908.8	18.4	1636	96	PFPMSPB9	D13356 P. falcipar
22	907.2	18.4	1636	96	PFPMSPB34	D13361 P. falcipar
23	907.2	18.4	1636	96	PFPMSPB44	D13363 P. falcipar
24	904.6	18.3	1636	96	PFPMSPB222	D13359 P. falcipar
25	785.6	15.9	3395	96	PFMSMSA1	X61930 P. falcipar
26	711.8	14.4	3518	96	PFAP190A	M35727 P. falcipar
27	711.8	14.4	3518	96	PFEP190G	Y00087 Plasmodium
28	691	14.0	1203	96	PFPMSP1	M64681 Plasmodium
29	655.8	13.3	1131	5	AF325919	AF325919 Plasmodium
30	623.6	12.6	1065	96	PFPMSPBPRO	L20092 Plasmodium
31	560.8	11.4	999	96	PFMSA1AC	M77715 P. falcipar
32	560.8	11.4	999	96	PFMSA1AF	M77718 P. falcipar
33	560.8	11.4	999	96	PFMSA1AG	M77719 P. falcipar
34	560.8	11.4	999	96	PFMSA1AH	M77725 P. falcipar
35	559.2	11.3	999	96	PFMSA1AA	M77713 P. falcipar
36	556.6	11.3	1133	5	AF061143	AF061143 Plasmodium
37	547.8	11.1	1080	5	AF266876	AF266876 Plasmodium
38	546.4	11.1	1956	96	PFMEZSA1D	Z35329 P. falcipar
39	544.8	11.0	1956	96	PFMEZSA1A	Z35326 P. falcipar
40	540.4	10.9	1141	5	AF061141	AF061141 Plasmodium
41	538.8	10.9	1142	5	AF061140	AF061140 Plasmodium
42	536.8	10.9	999	96	PFMSA1AH	M77720 P. falcipar
43	535.4	10.8	1959	96	PFMEZSA1C	Z35328 P. falcipar
44	528.2	10.7	1115	5	AF061146	AF061146 Plasmodium
45	527.8	10.7	1009	96	PFH3A	X52962 Plasmodium

OY	61	cagtcgtgaccacgaaatccctacagagctggttaagaacctggaagccttggaagat	120
Db	61	CAGTCGCGACCCACGAATCCTTACAGAGACTGGTTAAAGAACTGGAACTTTGGAGAAAT	120
OY	121	gcccctctaccggaatacaacgcctgttccgaagagaagaatggtgtctgaatgaaggagc	180
Db	121	GCCGCTCTTACCGGATATACCTCTGTTCCAGAGAAAGATGGTGCTGAATTAAGGGACG	180
OY	181	agtgcacgagccgttataacacgaacacccgggtctcaagggtctgtgtagcggtgac	240
Db	181	AGTGACACGGCCGTTAAACACGACACCCGGTTCTAAAGGGTCTGGCTTAGCGGTGGC	240
OY	241	tcgggtgggtctgtgacctctggggggttccgtccgctcccgcgagcgttgacatcaagt	300
Db	241	TCCGGTGGGTGTGAGCCCTGTGGGGGGTTCGTCGCTCCGGCGCAGCGTGGCATCAGGT	300
OY	301	ggctcaagtggcaagcggcggttccgynaacagtgcgaagaaccaatccatctgcgaactct	360
Db	301	GGCTCAGTGGCAACGGCGGGTTCGGGGAACGTGGAAGAACCAATCCATCTGACAACTCT	360
OY	361	aggcattccgagcgcgaagtctctcgcgcgaactaaacacgggtgaagaactatctctc	420
Db	361	AGCGATTCGCGAGCCAAAGTCTACGCGCACTTAACACCGAGTGAAGAACTATCTCTC	420
OY	421	actatcaagagcgctgaagtaccacaagctgttcgcacctcaactatatatgtctgacatg	480
Db	421	ACTATCAAGAGCGTGAAGTACCCACAATTTGTGCACCTCAGTATATATATGTGTGACACTG	480
OY	481	lgtgataacatctcatgtgctccaatatatctgattgacggttaccgaagaagtcacatgaactc	540
Db	481	TGTGATTAACATTTCAATGGCTTCAATATCTGATTTAGAGGTTAGCAAGAGATCAATGAACCTC	540
OY	541	ctgtacaagttgaaatttctactctcgtactgtgctaaagggccaacttgatagctttggcgc	600
Db	541	CTGTACAAGTTGAATTTCTACTTCTGCAGCTTGTCTTAAGGGCCAAACATGATATAGCATTGGCGCC	600

RESULT		ALIGNMENTS	
LOCUS	A92451		
DEFINITION	Sequence 2 from Patent WO9814583.		
ACCESSION	A92451		
VERSION	A92451.1		
KEYWORDS	GI:6741181		
SOURCE	unidentified.		
ORGANISM	unidentified		
REFERENCE	unclassified.		
AUTHORS	1 (bases 1 to 4940)		
TITLE	Pan, W. and Bujard, H.		
JOURNAL	METHOD FOR PRODUCING RECOMBINANTS INTENDED FOR USE IN A COMPLETE		
FEATURES	MALARIA ANTIGENE GP190/MSP1		
Source	Patent: WO 9814583-A 2 09-Apr-1998;		
	PAN WEIQING (DE); BUJARD HERMANN (DE)		
	Location/Qualifiers		
	1..4940		
	/organism="unidentified"		
	/db_xref="taxon:32644"		
BASE COUNT	1669 a 1157 c 1060 g 1054 t		
ORIGIN			
Query Match	100.0%; Score 4940; DB 9; Length 4940;		
Best Local Similarity	100.0%; Pred. No. 0;		
Matches 4940; Conservative 0; Mismatches 0; Indels 0; Gaps			
1	cgacgcgcatgataatcattcttcctccctcgttcattctcgttttttatcatcact 60		
1	CGACGCGCATGAATAATCATTCTTCCTCCTCGTTCATTCTCGTTTATCATCAACT 60		

Db	601	AAAGCATTTTGTCAAAATTCATTCCATTCAATTTTAAAGATCAGAGCCACAGAGTTGGACGATTGG	660
Qy	661	aagaagatggtgtcttcgcgatactcgaaagcctcttcgacaacatcaagacaatgctggaaag	720
Db	661	AAGAAGTGTGGCTTCGGATATGCCAAGCCCTCGACAAATCAAGACAAATGTGGAAAG	720
Qy	721	atggaagatatataaaagaataagaagaccatcggagaacattaaagactgtgtcgaa	780
Db	721	ATGGAAAGTTTATTTAAAGAAATAAACAAACCATGTGAGAACCTTAAACGATGTATCGAA	780
Qy	781	gaatccaaagaagaccatagacaataataagaatgcaaccaaggaaggaagaagaagaag	840
Db	781	GAATCCAAAAGACCATATGACCAAAAATTAAGAAATGCAACCAAGSAGSAAAGAAAGAAAG	840
Qy	841	ttgtaccagagcccgatagcagactgtccatctataaacaagcttgaagaagccataac	900
Db	841	TTGTACCAGGCGCCGTACGACCTGTCATCTATAACAAACAGCTTGAAGAAACCCATTAC	900
Qy	901	ctcaatcagcgtactcgtggagaagcgcgatagacaacctcaagaagaatgaaatatcaagaa	960
Db	901	CTCATCACCGTACTCGGAGAGCGCATATGACACCTCAAGAAAGAAATGAAATATCAAAAGAA	960
Qy	961	ctgtctgcgaagaatlaatgaaatlaagaatccctcgcgcgacaaactcttggaaacacct	1020
Db	961	CTGTCTCGACAAAGATTATGAAATTTAAACATCCCTCCGCCACCCAACTCTGGAAACACCCCT	1020
Qy	1021	aacaagctgtctgcgaacaagaacgaatagagagacagagaagaatgatacaagaagatc	1080
Db	1021	AACAAGCTGTCTGCACAGAAGCAAGAAAGATAGAGSACACAGAAAGAGATCAAAAGATTC	1080
Qy	1081	gccaaaacattaaagttcaacatagatctctccttctaacgtgatcccttgaagcttgaagtaac	1140
Db	1081	GCCAAAMCCATTAAAGTTCAACATAGATTCCTCTTAACTGATGCCCTTGAAGCTGGAGATAC	1140

[illegible]

Db	2221	AAGCAGGACAGCCTCCACAGGCCAGTGGCCGTTCAGAGGCTCAAGGCTCAAGTGGCTCA	2280
QY	2281	ccaccagctctctggaataacaagaccgggaatgtragaacactgacttaactgagag	2340
Db	2281	CCACCAGCTCTGTGAATTAACAAGACCGGAAGTGTGCGAAACTGGACTACTTGAGAG	2340
QY	2341	ctctatgagctccggaatacatctcatctgtgcacaaatatactcctgctctcgaagc	2400
Db	2341	CTCTATGAGTTCCTGATACATCTCTACATCTGGCAAAATATATCTCTGTCCTACAGC	2400
QY	2401	actatgaaacgagaatctttaaacsgtlacaagataaccaaggaagagagtaaacgtg	2460
Db	2401	ACTATGAACGAGGAGATTCTTAAACAGTACAGATPAACCAAGAGAGAGAGTAAATCG	2460
QY	2461	tcctctgtgtatccacgagccgcgtgtccatctcagaaacaaatcccgltatgcat	2520
Db	2461	TCCTCTGTGTATCCACGTGAGCGACTGTCTTCAATATCAACAACAATTCCTGGTATGAT	2520
QY	2521	tcctatgttcgaatagcctcaacaaatctctctctcgaactgtccatgagataatgagag	2580
Db	2521	TCTATGTTCCGATACCCCTCAACATTTCTCTCTCAACTTTCATGAGATATATGAGAG	2580
QY	2581	gagatggtctgcaacctgtataaaactaaagaacacgacaagattaagaacctctgag	2640
Db	2581	GAGATGGTCTCAACCGTATTAACATCAAAAGACAAGACAAATTAAGAACCTTCTGGAG	2640
QY	2641	gaagctaaagaagctctccacctctgtttaaactctctctccagctccatgcaacacgtg	2700
Db	2641	GAACTTAAGAAGGCTCCACCCTCTGTATAAACCTCTCTTCAGCTCCATGCAACCACTG	2700
QY	2701	tcctcacaacctcaagaacaagcccggaatgagcgctcaagaacacctcaacctcgacc	2760
Db	2701	TCTCTCAACCTCAAGACAAGCCCGGAAGTGAAGCGCTTAAGACAGCACTCTCACTCAAC	2760
QY	2761	aaccttaataactcaactgaaactgtttgagacatccgtctctcgcgaagaataagAAC	2820
Db	2761	AACCTTAATAACTCACTCAAGAAAGCTTTGGAACATCCTGTCTCTCGCAAGATTAAGAAC	2820
QY	2821	atctacaagaagacttatggaagaagaatcgccggaactcttaagaagaatgactgaa	2880
Db	2821	ATCTACCAAGACTTATTGGACAGAAATGCTCGGAACCTTTCAGAGAAAGATATCTGAA	2880
QY	2881	gaacgcgacacatctctaatacgaagacttcaactcaactctgtataactcaaacgcatgac	2940
Db	2881	GACAGCGACATTTCTATACGAGACTTCTCACTAACTTGTGAATCTTAAGCCGATAT	2940
QY	2941	atcaactctcttaaagatgaatcctaaacgtlaaagaactggaagagagatcaataagctg	3000
Db	2941	ATCAACTCTCTTAACGATGATCATTAACGTAACAACTCGAAGAGAGCATCAATTAAGCTG	3000
QY	3001	aagaagacactgcaactgagcttcgcgacctgacaaagaataaacctgaaactggagaga	3060
Db	3001	AAGAAGACACTGCAACAGACTTTCGACTGTACAACTATCAAACTGAACCTGAGAGA	3060
QY	3061	ctcttcgacaagaagaagacagtcgcgacgaagtataagatgacagataagaagttgactctg	3120
Db	3061	CTCTTCGACAAAGAAAGACAGTCCGCAAGTATAGATGCAATTAAGAAGTTGACTCTG	3120
QY	3121	ctcaagagacagcttgaagaacaaactcaactcactgaaacatccgaaacacgtlactgag	3180
Db	3121	CTCAAGGAGCGAGTTTGAAGCAAACTCACTCACTGAACAACTCGAAGACAGTACTGAG	3180
QY	3181	aactctccagtgtctctcaacaagaagaagagcccgagatcgcccgagacagagaacac	3240
Db	3181	AACCTTCACAGTTCCTTCAACAAGAAAGGAAGCGAATCGCGAAGACAGAGAACT	3240
QY	3241	ctggagaacaccgaagattctctcacaacactcaagagcctcgctaaagtatataatgggc	3300
Db	3241	CTGAGAGAACCAAGATTCTTCTCAAACTCAAAAGCCTGTGTAAGTATTATAATGAGC	3300
QY	3301	gagctctctctctgaagactctctccgagagagcatccagaccgagataaactaacgcc	3360

Db 3301 GAGTCTCTCTCTGAGAGACTCTCTCCGAGAGAGACATCCAGACCGAGATACACGCC 3360
 Qy 3361 agccctcgagaacttcaaggtcctgtcttaagctcgagaagctgaagcaaacctgaaac 3420
 Db 3361 ACCCTCGAGAACTTCAAGGTCTGTCTTAAGCTCGAAGGCAAGCTGAAAGCAACCTGAAAC 3420
 Qy 3421 ctgagaagaagaagctcagctacctctccttaagcgagctcaccacgtgtctcgagctc 3480
 Db 3421 CTGGAAGAAGAAGCTCAGCTACCTCTCTAGCGGAGCTCATACCTGATGCCCGAGCTC 3480
 Qy 3481 aaggaatcatttaagaacaagaactacacccggaataagcccaagcgagaaataatacagac 3540
 Db 3481 AAGGAAGTCAATTAAAGAACAGAACTACACCCGAATAGCCCAAGCGAATATATACAGAC 3540
 Qy 3541 gtagaataacgacactggaactcttacaagaagttcctgcccgaaggaagaaatgtccact 3600
 Db 3541 GTGAATTAACGCACTGGAACTTCTTACAGAAGTTCTGTCCGTAAGGAACACATGTGTGCACT 3600
 Qy 3601 gtgtgtctgaaactgtgtgtccgacacactgtagagcagttcacaactaagaagctgtcatc 3660
 Db 3601 GTGTGTCTGAATCTGGCTCCGACACACTGAGAGAGTCCAACTTAAGAAGCTGTGATCT 3660
 Qy 3661 actcatgtcggagccgagtgccacataataaccacatctcagaacgttcgaatgaagtc 3720
 Db 3661 ACTCATGTGCGGAGCCGAGTCCAAATACATACATCCAGACGTCGACGATGAGGTC 3720
 Qy 3721 gatgacgtcatcatgtgtgctcctcctcgagagcgaggaaggaactgaagatcactcgagc 3780
 Db 3721 GATGACGTCACTATGTGTGCTCTTCTTCCGCGAGAGCGAGAGAGACTAGCATACCTCGGC 3780
 Qy 3781 caggtgtgtcaaccggtgaggtgtgtcactccttcctgattgtataaactltgtccaaatc 3840
 Db 3781 CAGGTGTCAACCGGTGAGGCTGTCACTCCTTCCGATGTATACATCTGTCCAAATC 3840
 Qy 3841 gagaaagataaagaagtgctctctatctgaaaccttgcgagcggtctatagtgctctcaag 3900
 Db 3841 GAGAAAGAAATAGGAAGTGTCTATCTGAAACCTTGGCGGCGTGTATGTGCTCTCAAG 3900
 Qy 3901 aaacagctggaagaataaagctgtgaaccttcaatgttcaacgtgtgaagacatcttgaacagc 3960
 Db 3901 AAACAGCTGGAGAAATACCTGATGACCTTCACTGTCACGTAAGAGACATTTGTGAACAGC 3960
 Qy 3961 cgctttaaagaagaagaatttcaagaacgtctgtgagagcgacttgatctccataaa 4020
 Db 3961 CGCTTAAATGAAGAGAAATTTCAAGAACGCTTGGAGAGCGACTTGATCTCCATATAA 4020
 Qy 4021 gacccgacactccttaactcaagtgctgaagagccatcaagttcctcaataaagaag 4080
 Db 4021 GACCTGACCTCTCTTACTACTGCTGTCAAGGACCCATACAAAGTCTCTCAATAAAGAAG 4080
 Qy 4081 aggaataaattctgtctagttacaactatatacaagagctcatcgacacccatatacat 4140
 Db 4081 AGGATAAATTTCTGTCTAGTTACAACTATATCAAGAGCTCATGACACCCGATATCAT 4140
 Qy 4141 ttgcgttaatgctgtcgtgggtattacaagaatcctgaagcgaaataataaagtctgaacct 4200
 Db 4141 TTGCGTAATGATGTCTGTGGGTATTACAAGATCTGAGCGAAATAATACAAAGCTGACCTT 4200
 Qy 4201 gactctatataaagaatatacaagaataagcaagcgagaagaaataataatctgccttc 4260
 Db 4201 GACTCTATTAATAAAGTATATACAGATTAAGCAAGCGAAGAAATGAATAATGTGCCCTTC 4260
 Qy 4261 ctgaataacatcgaaacctgtfacaagaacgtgaagcagaacaaatgacctcttgtaatt 4320
 Db 4261 CTGAATAACATGAAACCTGTACAAAGAGAGTAAGCAACAAATGACACTCTTCTGTAATT 4320
 Qy 4321 caactgagggccaaggtccctcaactatacttaagaaagcaatgtgtgaagttaaactc 4380
 Db 4321 CACCTGAGAGGCCAAGGTCTCAACTATACTTACGAAGAAGCAATGTGGAAGTAAATC 4380
 Qy 4381 aaggaagctgaactactcaaaaacatccaagaacagctgtgcgagattcaagaaaaaataac 4440
 Db 4381 AAGGAGCTGAACCTCACTCAAAAACATCCAAAGCAAGCTGCGAGATTTCAGAAATAATAC 4440

Qy 4441 aatttcgtcgaatttcgaacactgtctctaccgattataaccacaacatctcctgaagc 4500
 Db 4441 AATTTCGTGGAATTGCACACTGTCTTACCGATTATATACCAACATATCTCTGACCAAG 4500
 Qy 4501 ttctgtccactgcatgggtgtctcgaaaaaccccgcaaaacagtgctgagcaatcgtctc 4560
 Db 4501 TTCTGTCTCAGTGGCATGGTGTTCGAAMAACCTCGCAAAACAGTGTGAGCAATCTGCTC 4560
 Qy 4561 gacggcaacctgcagggcatgtctgaacatctcccaagcacaatgtcgtgaagaacagtc 4620
 Db 4561 GACGGCAACCTCGACGGCATGTCTGAACATCTCCACAGCAACATGGCTGAAGAAACAGTC 4620
 Qy 4621 ccccaagaatgcgctgtgttcagagcatctgcgagcgagcgaggaagtgcaagtgctcc 4680
 Db 4621 CCCCAGAATAGCGCGCTGTTTCAGCGCATCTGGACGAGCGGAAGATGCAAGTGTCTCTG 4680
 Qy 4681 aactacaacaagaagaagataagtgctgtgagaaccccaacccctacactgtcaatgaaac 4740
 Db 4681 AACTACAACAAGAGAGAGATAGTGCCTGAGAACCCAAACCTTACCTGAATGAAAAAC 4740
 Qy 4741 aatggcgggtgtgacgacgatgtctaaatgtcacccggaggaagacagcgctctaaagaa 4800
 Db 4741 AATGGCGGCTGTGACGCCGATCTAAATGCACCGAGGAAGACAGCGCTCTTAACGGAAG 4800
 Qy 4801 aaatcacatgagtgatgtactaagcccgactctatacactcttcgaaggaatttttgc 4860
 Db 4801 AAATTCACATGGAGATGTACTAAGCCGACATCTCTTACCACTTTTGACGAGATTTTTC 4860
 Qy 4861 tccagctcaatttccctgggacatcctcctcctgtcatcctcaatgctgactcagc 4920
 Db 4861 TCCAGCTCAATTTCTCGGCACTCTCTTCCGTGATCCATGATCATCTGCTGTAACAGC 4920
 Qy 4921 ttcatctaataagatcagtg 4940
 Db 4921 TTCATCTAATAGATGATG 4940

RESULT 2
 CVEL31294
 LOCUS CVEL31294 4940 bp DNA SYN 24-FEB-1999
 DEFINITION Cloning vector pBSK*-MSP-1s/FCB.
 ACCESSION AJ31294
 VERSION AJ31294.1 GI:4455908
 KEYWORDS Cloning vector; major surface protein-1; map-1 gene.
 SOURCE Cloning vector pBSK*-MSP-1s/FCB.
 ORGANISM Artificial sequence: vectors.
 REFERENCE 1 (bases 1 to 4940)
 AUTHORS Bujard, H.
 TITLE Direct Submission
 JOURNAL Submitted (14-DEC-1998) Bujard H., ZMBH - Zentrum fuer Molekulare Biologie, Universitaet Heidelberg, Im Neuenheimer Feld 282, 69120 Heidelberg, BW, GERMANY
 REFERENCE 2 (bases 1 to 4940)
 AUTHORS Pan, W., Rayov, E., Tolle, R., Frank, R., Mosbach, R., Turbachova, I. and Bujard, H.
 TITLE Vaccine candidate MSP-1 from Plasmodium falciparum: a redesigned 4917 bp polynucleotide enables synthesis and isolation of full-length protein from Escherichia coli and mammalian cells
 JOURNAL Nucleic Acids Res. 27 (4), 1094-1103 (1999)
 MEDLINE 99128299
 FEATURES
 source location/Qualifiers
 1..4940
 /organism="Cloning vector pBSK*-MSP-1s/FCB"
 /db_xref="taxon:85834"
 /focus
 1..4940
 /organism="Plasmodium falciparum"
 /strain="FCB-1"
 /db_xref="taxon:5833"
 10..4929
 /gene="msp-1"

```

sid_peptide      10. .69
                  /gene="msp-1"
                  /gene="msp-1"
                  /function="surface protein"
                  /codon_start=1
                  /evidence=experimental
                  /product="major surface protein-1"
                  /protein_id="CA836903.1"
                  /db_xref="GI:4455909"
                  /translation="MKILPFLGSLPFIIMQCVTHSEYQGLVAKSLLEALPDALVTGS
LPQEKERKVLNMGTSQTRVTTSTPSKSLKSTVSDSGSGSVASGSGVSLGASGVSAS
GGSQSRRTNPSDNSDSDAKSTADLKHRRNRNLVLTIKELYQLPDLTHHMLTLDN
IHGRKYLIDYEIEINELLYKLNFFYDLKRAKLDVNCANDYQCOIPFNLIKANELDYLK
KLVEGRKPLDNIKDNYGKMEDEYIKRKRTIENINELIEESKRTIDKRNKATKEEEK
KLVAOYDRLITYNKOLEEHNLIISYLEKRTIDPLKPNELIELDKINIEKNPPANG
NTPLDLDLKKKTIIEHKERKELIAKTIKFNIDSLFNNELIETIYLRKKNIDISAKV
ETKESTENPEYNGSVYTPLEYNIDNNALNELNSFGDILNFFDYTKESKITIYDNEK
KFIENIKETIKTEKKIIESDKSSTEDSKSLNDITKYEKLNIEYDSKTRNNIIDLIN
FEKMGKRSYKVEKELIHNHTFASYESKNLEKLTIKALYMEYISLRNIVKRELYK
YKNLISKIENIEITLIVENIKKDEOLEEKITKDEKNPDKLILEVSIYVQOKVLL
MKNLDELKTQOLIKANVELKHNHIVPYSRQONKOEYVLIIVLKIEDIKLVMPKVE
SLINEEKNKIKTEGQSDNSPREGETGCATTRPGQASALEGDSVQAQADOKQA
OPPPVYPPAKAOPIPPAPVNNKTEVNSKLDLELYLFNTSTYICHYLTIVSHT
MNEKLIKQYKITEEESKSLSCDPLDLFNLIONNIPVMSFPLSNYSKILTSIET
KEWCMCNLKLKDNDIKNLLEAKKAVSTSVKTLSSMQGLSTPOKRPVSNADTSS
HSTNLNSLKLLENILSLGNKNKIYOELLIOKSESNFEKILDSDFEYSESTPNYK
SKADDINSLNDESKRRKLEDDINKIKTQLDLSPIYKRYKLTLEPLDKKVTGYKYM
QIKRTILKLEQLESKLSLNNPKHIVLONSFYFNRRKKKELAIETENLENTLIKLYH
KGLKYYNGESSPLKLTLESISTOEDVASLENKVLKLEGLKLDNLEKRLSTL
SSGLHMLIAELKEVIAKNKNTGNSPSSNNIDVNNALLESYKRLPEGTDVATVSEGS
DTELOSQPKRPASTHVGAESNTITTSQNDVDEVDVITVIFEGSEEDVYDLQAVYG
EAVPVSVIDNLISKIEMEYEVLYKPLAGYRSRLKOLENNMTFNNVNDIILSRN
KREKFNKLVESDPLIPYKDLTSSNNVYVDPKPLINKERDKFLSSYNTIKDSITDINF
ANLYQGYKTLSEKYYSDLSIDISKIYINDKGENEMKIPLIENIETIKYINDKIDLIV
IHEAKVLNNTTEKSNVEYKIKELNTYKTLTODLKLADKKNNNEVGLADLSTVHNHL
LTLEFLSTGVNFENLAKTVLISNLDGNLQMLNLSOHCVRKQCPUNSGCPSRHLDERE
CKCLINTKQEGDKCENPNPTCNENNGGCAADAKTEEDSGSNGKLTTCCTCPDPSYP
LQFQIFCSSNFIQIPIFLILMLITLYSFI"
BASE COUNT      1669 a 1156 c 1060 g 1055 t
ORIGIN
Query Match      100.0%; Score 4938.4; DB 56; Length 4940;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4939; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 cgcacgcatgaaatcatttcttctctctgttcaatttctgtttttatcatcaataact 60
Db 1 CGCAGCGGTATGAAAAATCATTTCTCTCTCTGTTCAATTTCTGTTTATCATCAATACT 60
OY 61 cagtcgctgaccaccgaatctcatcaagaagctgtgttaagaacgtggaagctttgaagat 120
Db 61 CAGTCGCGTACCCACGAATCTCATCAGGAGCTGTGTAAAGAACTGGAAGCTTTTGGAAGAT 120
OY 121 gcgcctcttaccggaataagcctgttccagaagaagaaatcgtgtcgtgaagaaggagc 180
Db 121 GCCGCTTACCAGGATACAGCCTGTTCAGAAAGGAAGAAGATGCTGCTGAATGAAGGAGC 180
OY 181 agtggcagcgctgttacaacacgacacaccgcggttctaaaggctgtgtgctagcgttgc 240
Db 181 AGTGCAAGGCGCGTTACAAACACAGACACCCTGTTCTAAAGGCGTGTGGCTAGCGGTGC 240
OY 241 tcgcgttggtctgtgacctctgggggttccgtgcctcccgcgcgcaagcgttgtaagagt 300
Db 241 TCCGGTGGGCTGTGTGGCTCTGTGGGGGTTCCTGCTCCCTCCGGCGGACGCTGGAAGT 300
OY 301 ggtcactggtgcaaacgctgttccggaagaagtcgaagaacaaatccactgcaaacactct 360
Db 301 GGCTCAGTGGCAAGCGCGGCTTCGGGAGAACGTGGAAGAACCAATTCATCTGACAACCTCT 360
OY 361 agcgattcgaagcccaagctctctacgcgcgaacctcaagcaccgagtggaataactctctc 420
Db 361

```

Dd	361	AGCGATTCCGAGCGCCAAAGTCCTACAGCCGACCTCAGACCCGAGTGAGAAATATCTCCTC	420
Qy	421	actatacaagagcgtgaagtlaccacaagctgttcgcacctacaatcalatagtgcactg	480
Dd	421	ACTATCAAGAGAGCTGGAAGTACCCACAGTTTGTGACCTCACTATATCTGTGCACACTG	480
Qy	481	tggtataacattca tggcttcctaataatctgattgacggtttagaagagatcaatgaactc	540
Dd	481	TGTGATTAACATTTCATGSCCTTCAATATNTGATTTGACGGTTTACGAGAGATCAATGAACCTC	540
Qy	541	ctgtatacaagttgaa tttcttactcttcgcacttgctaaaggtccaaactgaatgaagctgttggcc	600
Dd	541	CTGTATCAAGTGAATTTTCTACTCTTCGACTTGCTAAAGGCCAAACTGAATGACCTTTGGCCCC	600
Qy	601	aatgactatgttcaaat ttcattccattcaatttgaagaatcaagggccaacagatttgaactg	660
Dd	601	AATGACTATTTGTCAAAATTCATTCCATTCTTGAAGATCAGAGCCACAGAGTTGGACGTATTG	660
Qy	661	aagaagttggtcttcggtat tgcgaagcctctgcacaatcaaggaacatgttgggaaag	720
Dd	661	AAGAAGTTGGTCTTTCGGATFATTCGAGAGCCCTCTGCACACATCAAGACAAATGTGGGAAG	720
Qy	721	atggaagatatataa aagaataaagaagcacatcgagaacattaaegagctgtacgaa	780
Dd	721	ATGGAAGTATTATTTAAAGAAATAGAAAGCCTATGAAACATTATACGAGCTGTATCGAA	780
Qy	781	gaattccaaaagaccataga caaaataaagaatgagaccagaagggagaagaagaagaag	840
Dd	781	GAATCCAAAAGACCATTAGACAAAATTAAGAAATGCAACCAAGAGAGAAAGAAAGAGAG	840
Qy	841	ttgttaacagggccag tlaagcagccttgcattatataaacaacagcttgaagaagcccatlac	900
Dd	841	TTGTATACGAGGCCAGTACGAGCTGTCCATCTATTAACAAACAGCTTGAGAAAGCCCATAC	900
Qy	901	ctcatcagcgtcatctg gagaagcgcatagacacccctcaagaagaatggaataatacaagaa	960
Dd	901	CTCATCACCGTACGTGGAAGGCCCATAGACCCCTCAAGAAATGAATATCTCAAGAA	960
Qy	961	ctgtgcagacaagaat taaatgaatccctccggccaagccaactcttgggaacacccct	1020
Dd	961	CTGTCTGCACAGATTATATGAATTTAAGAAATCTCCGCCACCACTCTTGAGAACCCCT	1020
Qy	1021	aacacgcgtgcgtgga caagaacaagaagatagagagcaagagaagaagatcaagaagatc	1080
Dd	1021	AACACGCGTGTGCGACAGAACAGAAAGATAGAGACACAGAAAGAGATCAACAAAGATC	1080
Qy	1081	ggcaaaacattaa gttcaacatagatctctcttactgatactcccttgaagcttgaagtac	1140
Dd	1081	GCCAAAACCATTAAGTTCAACATATAGATTTCTCTTACTTACGATCCCTTGAGCTGGAATAC	1140
Qy	1141	tactttgagagagaaga aatagaatatagacatctccgcaagaatcgagacaagaagatca	1200
Dd	1141	TACTTTGAGAGGAAGATTAAGATTAAGACATCTCCGCAAAAGTGAAGCAAGAAATCA	1200
Qy	1201	accgaacctaatgtata tcccgaatgtgtgaacgttaccctctgtctataacagatcaac	1260
Dd	1201	ACCGACCTATATGAAATATCCCAATFGGTGTGAGCTACCCCTGTCTTATATACGATATCAAC	1260
Qy	1261	aacgctctcaagcagctcaatagcttcggtgactgtatataacccctcgatatacgaa	1320
Dd	1261	AACGCTCTCAACGAGCTCATATACCTTCGTGATCTTATTAACCCCTTGATATACGAAA	1320
Qy	1321	gaaccccttaagaa tactctacacagcaatgagagaagaattctatacagaagaatcag	1380
Dd	1321	GAACCCCTTAAGATATCTACACAGCAATGAGAGAAAGAAAGTTATACAGAAATTCAG	1380
Qy	1381	gagaagaatcaaaat tggagaagaagaattgagagtgagacaagaagaaagttlacaagaagccg	1440
Dd	1381	GAGAAGATCAAAATTTGAGAGAAAGAAATTTGAGAGTGACAAAGAAAGTTACGAGAGACCC	1440
Qy	1441	agcaaaagttcaacgat taccataaagatlbgaagcgtgcgtgaacgagatctatgat	1500
Dd	1441	AGCAAAAGTCTAAACGATATACCTAAAGATATGAAAGACCTCTCAAGAGATCTATAT	1500

QY	1501	tcgaatctcaacataataatcatcgccctcgaccgaacttcgagaaatgatatgagaaacggtac	1500
Db	1501	TCGAAATTCACACATTAACATCGACCTCGACCACCTTCGAGAAATATGAGGAAACGGGAC	1500
QY	1561	tcttacaagaatgagaaacatgacacacccaataaccttgcatccctatgagaatctctaa	1620
Db	1561	TCCTTACAAAGTGGAGAACTGACACACCATTAATACCTTTGCATCTCTATGAGATTTCTAAG	1620
QY	1621	cataatcttgagaaagctcaccaagctcttaagtatatgagagactattctctcggaac	1680
Db	1621	CATATCTTGAAGAGCTCACCAAGCTCTTAAGTATATGAGAGACTATTTCTCGGAGAC	1680
QY	1681	atgtcttgagaaagaacctaagaatctacaagaatctcataagtaagatcgaaacagag	1740
Db	1681	ATGTCTTGGAGAAAGACTAAGATTTACAGAAATCTCATATAGTAAATCGAAACAGAG	1740
QY	1741	atcgagacgcttgcttgagaacaltaaagaagatgaaacagtgcttgagaaagaatt	1800
Db	1741	ATCGAGACGCTTGCTTGAGAACATTAAAGAAGATGAAAGAACTGTTTGAGAACAGATTT	1800
QY	1801	acaaagaacgaagaataaaccagaatgagaagaatccctggagagctccgatattgttaagtc	1860
Db	1801	ACAAAGAACGAAATTAACCCAGATGAGAAAGATCCTGGAGGCTCCGATATTGTTAAGTTC	1860
QY	1861	caagtgcagaagtgctctccctcctcgaacaaagatttgaaactcaagaagatcacaatcat	1920
Db	1861	CAGTGCAGAGGAGCTCCTCCATGAGAACAGATGTGATGAACTCAAGACCTCACTCATTT	1920
QY	1921	ctgagaagaacttgagattaaacaataataactctgctgacgaatagttataagcagagagat	1980
Db	1921	CTGAGAGAACGTTGGAGTTAAACATTAATTAATACATGTGCCAATAGTTTATTAACGAGAGAT	1980
QY	1981	aagaagaacacatactactaccctcactgcgtacccaagaagaagatgatgacaacatgaagtgttc	2040
Db	1981	AAGCAGAGAACCATTAATCACTCACTCACTCAAGAAAGAAATGACAAACTGAAGTGTTC	2040
QY	2041	atgcgccaaagtctgagagccttgatccaacgagaagaagaagaacattaaactgaaagagac	2100
Db	2041	ATGCGCCAAAGTCGAGACCTCGATCAACGAGAGAGAAAGAAACATTAAACTGGAAGGACAG	2100
QY	2101	tcgataatactccgagaccttcacagaagaagagataaaccgagcaggtctaccacaacgcc	2160
Db	2101	TCGATTAATACCTCCAGGCTTCCACAGAGAGAGATTAACCGGACAGGCTTACCAACCAAGCCC	2160
QY	2161	ggacaacaagccgggtctcgaagctcgcgaaggcgaatagcggtcgaagctcaagcaagaagcag	2220
Db	2161	GGACAAACAGGCGCGGTTCAGCTCTCGAAGGCGATAGGCGTCAAGCTCAAGCACAAGAGCAG	2220
QY	2221	aagaagaacagagctccagatggccagatggccggtctccaagaggtcctaaagctcaagtgctaca	2280
Db	2221	AAGCAGGACACAGCCTCCAGTGGCCAGTGGCCGTTCCAGAGGCTTAAGGCTCAAGTGCCTACA	2280
QY	2281	ccacagagctccttgaaataaacaagaccggaatgtcgaagaaactgacataccttgagaag	2340
Db	2281	CCACACAGCTCCTGTGAATATACAGAGACCGAAGATGTGACGAACCTGACTTACCTTGAGAG	2340
QY	2341	ctctatgagttcttgaaatatacctctacatctgcgcacaatatataccttgctctctccacgc	2400
Db	2341	CTCTATATGAGTTCTTGAAATATACCTCTACATCTCCACAAATATATCTCTGCTCTCCAGGC	2400
QY	2401	actatgaacgagaagaattctctaaacgatacaagataaaccagaagaagagaagtaaatg	2460
Db	2401	ACTATGACAGAGAGATTCTTTAANCAGTACAAAGATTAACCAAGGAAGAGAGATTAACCTG	2460
QY	2461	tccctctgtgataccactggaacccgctgcttcaatctacccaagaacaacatcccggttatgat	2520
Db	2461	TCCCTCTGTGATCCACTGGACCGCTGCTTCAATATTCGAGAACAACTCCGCTTATGAT	2520
QY	2521	tctatgtctgataagctccaacaattctctctctccaactgltcaatgagagatatatgagaag	2580
Db	2521	TCCTATGTTTCATAGGCTCCAACAATTTCTCTCTCTCCAACTGTTCAATGGAATTAATGAGAG	2580

QY	2581	ggaatgagctcggaaccttataaactcgaagaacagacaagatataaactctcgag	2640
Db	2581	GAGATGGTGTGCAACCTGTATTAACCTCAAAAGCAACGACAAGATTAAAGACCTTGTGAG	2640
QY	2641	gaagctaagaagtcctccacacctctttaaactctctctccagctccatgcaacacatg	2700
Db	2641	GAAAGCTAAGAAGCTTCCACCTCTTTAAACTCTCTTCCAGTCCATTCGCAACCACTG	2700
QY	2701	tctctacaacctcaagacaagcccggaatgagcgctaaagacacctctcaatcgacc	2760
Db	2701	TCTCTCACACCTCAAGACAAGCCGCAAGTGGCGCTTAACGACGACACTCTCACTCGACC	2760
QY	2761	aaccttaataactcctaaactcttgaagaactccctgctcttcggaagaataagaac	2820
Db	2761	AACCTTAATTAATCACTAACAAGTGTTTAGAACATCTGTCTTCTGGCAAAATTAAGAC	2820
QY	2821	atctaccaaagaactatctgacagaaatcgcgtccggaactctacgaagaatactgaa	2880
Db	2821	ATCTACCAAGAAGACTTATGGACAGAAATCGTCCGAGACTTCTACGACAAATACTGAA	2880
QY	2881	gaacagacaacatctcatalaagagagctcaactaactcgttgaatctaaagccgatg	2940
Db	2881	GACACGACACACTTCTATAAGAGAGGCTTCACTTAACCTCGTGAATCTAAAGCCGATAT	2940
QY	2941	atcaactctcttaagatgaatcttaaagtgtagaagctcggaagaggacataaagtctg	3000
Db	2941	ATCAACTCTCTTAACGATGAATCTTAAGGTAAAGAGCTGGAAGAGGACATCAATTAAGCTG	3000
QY	3001	aagaagacactgcaactagcttcgacctgtgcaacaagtacaactgaacttgagaga	3060
Db	3001	AAGAAGACACTGCAACTAGCTGAGCTTGCACTGTATCAACAAGTACAAACTGAACCTGGAGGA	3060
QY	3061	ctcttcgaacaagaagaacagtcgcgaagatataagatgcagatcaagaagtgcactctg	3120
Db	3061	CTCTTCGACAAGANAMACAGTCCGCAAGTATAAGATGACGATCAGAAAGTTCACTCTG	3120
QY	3121	ctcaagagcgagcttgaagaacaaactcaactcaactgaaacaaatccgaaacacgactcgag	3180
Db	3121	CTCAAGAGCGAGCTTGAAGAACAACCTCAACTCACTGAACAATCCGAAACACGTACTCGAG	3180
QY	3181	aactctcaagtcctctcaacaagaagaagaagccgagatcgcgcgagacagagaacact	3240
Db	3181	AACCTCTCAGTCTTCTTAACACAGAAGAAGAGCGGAGATCCGCCGACACGAGAACCT	3240
QY	3241	ctggaagaacaccagaatctctccaacaactacaagaagccctcgtaagtatatataatgac	3300
Db	3241	CTGGAGAACAACCAAGTCTTCTCAAACTATCAAAAGGCTCTGTCACATTTATTAATGGC	3300
QY	3301	gagcttctctctcgtgaagacctctctccgagaagagcatccaagccggaataactaagcc	3360
Db	3301	GAGCTTCTCTCTGTAAGACTCTCTCCAGAGAGACATCCAGACCGAGATTAACCTAGCC	3360
QY	3361	agccctcgagaactcgaagtcctgtctgaagtcgaagcgaagctggaagaacacctgaac	3420
Db	3361	AGCCCTCGAGAACTTGAAGTCTCTGTCTAAGCTCGAAGGCAAGCTGTAAGACACACTGAAC	3420
QY	3421	ctggaagaagaagctcaagctacacctctctagcggagctgcaatcaactgcatcgccgagctc	3480
Db	3421	CTGGAGAAGAAGAACTCAAGTCACTCTCTAGCGGACTCATCACTGATGTCGGAGCTC	3480
QY	3481	aagaagatcaltaaagaacaagaactacaacgagcaatagcccaagcgaagaataatacagac	3540
Db	3481	AAGGAATCATTTAAGANAMCAAGAACTACACCGCAATAGCCCAACGGAATTAATACAGAC	3540
QY	3541	gtgaaataacgcactcggaaatcttacaagaagtctctgcccctgaagaacaagatgtgcacact	3600
Db	3541	GTAATTAACGCACTGGAATCTTCAAGAAGTTCCTGCTTGAAGAACACGATGTGCCACT	3600
QY	3601	gtgagtcctggaatcctcgagctccgaacaacttggaagatctcaaacctaaagaagctcgacatc	3660
Db	3601	GTCGTCCTTGATCTGGCTCCGACACACTGGAGCAGATGTCAACTTAAGAACCTCTGCATCT	3660
QY	3661	actcatgtccggaagccgaatccaatacaattacccaacttcgaagacgtcgagatgaggttc	3720

[illegible]

Db	4741	AAAGCGGGGTGTGACGCCGATCTTAAATGTACACCGAGAAAGACAGCGGCTCTAACGGAAG	4800
Oy	4801	aaatcacatgcagtgltactaagaccgactctcatcactcttcgacggagatltttgc	4860
Db	4801	AAAATCACATGCGAGTAGTAAAGCCGACACTTCTATCACACTTTCGACGGAGTTTTFGC	4860
Oy	4861	tccagctcaatttcctcgtggatctctctctgcgcgatacctcatatgctatctgtacgc	4920
Db	4861	TCCACCTCTAAATTTCTTGGGCACTTCTTCCTCCGTCGATCTCATCTGATCTGATCTGACAGC	4920
Oy	4921	ttcactaataagatcatg	4940
Db	4921	TTCACTTAATAGATCATG	4940
RESULT	3		
LOCUS	PFANT195	5917 bp	mRNA
DEFINITION			Plasmodium falciparum mRNA for major surface antigens precursor (P195).
ACCESSION	X02919		
VERSION	X02919.1	GI:9864	
KEYWORDS			antigen; direct repeat; signal peptide; surface antigen.
SOURCE			malaria parasite P. falciparum.
ORGANISM			Plasmodium falciparum
REFERENCE			Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS			1 (bases 1 to 5917) Holder, A.A., Lockyer, M.J., Odink, K.G., Sandhu, J.S., Riveros-Moreno, Y., Nicholson, S.C., Hillman, X., Davey, L.S., Tizard, M.L.V., Schwarz, R.T. and Freeman, R.R. Primary structure of the precursor to the three major surface antigens of Plasmodium falciparum merozoites
TITLE			Nature 317 (6034), 270-273 (1985)
JOURNAL			86014355
MEDLINE			2 (bases 1 to 5917)
REFERENCE			Holder, A.A.
AUTHORS			Direct Submission
TITLE			Submitted (06-MAR-1991)
JOURNAL			Holder A.A., National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK
COMMENT			Update of published sequence
REMARK			Data kindly reviewed (06-MAR-1986) by R.R. Freeman.
FEATRES			Location/Qualifiers
Source			1. 5917 /organism="Plasmodium falciparum" /db_xref="taxon:5833" 418. 474 /product="put. signal peptide (aa -19 to -1)" 418. 5337 /note="P195 precursor" /codon_start=1 /protein_id="CAA26676.1" /db_xref="GI:9865" /db_xref="SWISS-PROT:P04933" /translation="MKILIFLCSFLPFIINTQCVTHESYDELVLKLEALEDAVLTVYS LFOKEKWLNGESTGTAATVTPSGSKDSVAGSGSVASGSAVSGSVASGSAV GSGSRRTPNDSNDSDAKSYADLKHRYANVLTITKELVYPOLPLTNMLTCDN IHGKYLIDGEEINELLYLNPFYDLRLALNLVNCANDQIPFNLTIRANELDVLK KLVGKRLPLDNIKDNVGMKMDYIKKKKTKTTRKNIIDSKKTIKDKNATKREKQ KLVAQADLSYKQKLEBAHLISVLEKRIIDTLKKNENIKELDKINIKRPPANG NTPPTLLDKNKKIEHEKEIKETAKTIFNPDS.LFTDPLEVLYREKNNIDISAVV ETPSTEPNEYPNGVTPPLSYNDINNALNLSFGDLPNDYKPEPSKNNTYDNERK KFLIEIKREKIKIEKKIESDKSYEDRSKSLNDITTEKELKLNIDYSGFNNDLNN PEKMKGSYKVKELTHHNTFASYSKSHNLKLTALAKMEDYSLRNIVKELKYL KYNLSIKLEINLEIIVENIKKDEORLEKKTREKNDPDKILVDSIVVQKQVLL MKNLDLKTOLLIKNVLKLNHVPVSKQENKOEYVLTIVLKEIDKLAKVPKYE SLINEEKNITKTEGSDNSESTGEITGQATTKPGQAGSALEGSVVAQAKQKQA QNPVLPVPEAKAQPVPAPVNNKTNVSLADYLEKLYEFLNLSYICHKYLIVSHST OMPELILQYRITKEESKLSLSCDPLDLNFIQNNI PVMSFDSLNNLSOLFWEIYE KEMVCNLYKLKDNIILSKNKNKYOLLIQKSENFYKELIKSDPFNSFPNPKV HSTLNLSIKLFENIISLKNKNKYOLLIQKSENFYKELIKSDPFNSFPNPKV SKADINSLNDESKRRKLLEDINKLTKTLDLSPILNKNYKILBERLDDKKTGVKTKM QIKRVLTLKQLESKLSLNPKKRVAVLQNSVFENKKEAEIAETENLLENKRLILKHY GVLKYLTLKQLESKLSLNPKKRVAVLQNSVFENKKEAEIAETENLLENKRLILKHY

SSGLHLLIAELKEVIRKNTGNSPSENNTDVNNALESYKFLPEGTDAIVASEGS
DLEQSORPKPASTHVGAESNTITTSQNVDDVDVVIIVPIGESEEDDDGOVYTG
EAVPSPVIDNITLSKILENEVYLKPLAGVRSKLENNMTNVNKKDILNREN
KRENNKYLESDLIPIKDOTSNNYVKPKYKLNKKRKEPSSNNYIKDSIDTEPINF
ANDVGYKILSEKIKSDSDSIKTYINDQGENEYLPPLNNIELYKVNKTIDLFV
IHLEKVLNNTYTERKSNVEKIKELNLTIDOKLADFKNNNEVIAIDSTDYNNHL
LTKFISTGVNENLAKTVLSNLDGNLQGLMNI SOHCVKKQCPNSGCFRILDEREE
CKCLINTYKOEQCEVNEPPTCENNNGGCDADAKCTEEDSGSNGKKTCECTKPDSP
LEDGIFCSSNNLGISFLLMLILYSFI"
475..5334
mat_peptide
/product="region coding for 83K, 42K and 19K major surface
antigens (aa 1-1628)"
631..728
repeat_region
/note="alternating repeated SVA-SGG (aa 72-104)"
misc_feature
763..765
/note="pot. glycosylation site"
misc_feature
1219..1221
/note="pot. glycosylation site"
misc_feature
2707..2709
/note="pot. glycosylation site"
misc_feature
2719..2721
/note="pot. glycosylation site"
misc_feature
2764..2766
/note="pot. glycosylation site"
misc_feature
2947..2949
/note="pot. glycosylation site"
misc_feature
3175..3177
/note="pot. glycosylation site"
misc_feature
3307..3309
/note="pot. glycosylation site"
misc_feature
3589..3591
/note="pot. glycosylation site"
misc_feature
3910..3912
/note="pot. glycosylation site"
misc_feature
3937..3939
/note="pot. glycosylation site"
misc_feature
4750..4752
/note="pot. glycosylation site"
misc_feature
/note="pot. glycosylation site"
BASE COUNT 2644 a 636 c 727 g 1910 t
ORIGIN

Query Match 57.0%: Score 2816.4; DB 96; Length 5917;
Best Local Similarity 73.2%: Pred. No. 0;
Matches 3609; Conservative 0; Mismatches 1321; Indels 0; Gaps 0;
QY 10 atgaaatcaatttcctcctcgttcattcgtttttatcaatcaatcagtcgctg 69
DB 418 ATGAAATCATATATCTTTTATCTTTTATCTTTTATTAATAACACAAATGTGA 477
QY 70 acccaagaatccatcagaagcgtgttaagaacttgaaagcttgaagaatgcgtcctt 129
DB 478 ACACATGAAGAGTATCAAGAACTTGCAAAAACCTGAAGCTTTAGAAAGTCAGATTG 537
QY 130 accgaatacagcctgttcagaagaagatgctgaatgaagaagcagtgagcag 189
DB 538 ACAAGTTATAGTTTATTCAAAAGGAAAAATGCTTAATAATGAAGACAAATGGAACA 597
QY 190 gccgttaacaacgaacaccggttctaaaggctcgtgctagcggctggtccgtg 249
DB 598 GCTGTACACACTAGTATGACTGCTTCAAAAGGGTTCAAGTTGCTTCAAGTTCAGTGC 657
QY 250 tctgtgacctcgggggttcgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 309
DB 658 TCAGTGTCTTCAGGTGCTTCAGTGTCTTCAGTGTCTTCAGTGTCTTCAGTGTCTTCAGT 717
QY 310 gcaagcggcgttcgc 369
DB 718 GCTTCAGGTGCTTCAGTGTCTTCAGTGTCTTCAGTGTCTTCAGTGTCTTCAGTGTCTTCAG 777
QY 370 gacgcgaagtcctcgaagcgaactcaagcgaagtgagaatactctcctcaatcaag 429
DB 778 GATGCTAAATCTTACGCTGATTTAAACACAGAGTACGAAATTAATTAATTAATTAATTA 837

QY 430 gacgtgaagtlaccocacagttgttcgacctcctcactcaatcatatcgtgacctgtgtataac 489
DB 838 GAACATAAATATCTTAACCTTTTGATTTAACTAATCATATGTTAACTTTGTGTGATAT 897
QY 490 attcatgcttccaataltcgtatgcggctacgaagaatcaatgaactcctgtataag 549
DB 898 ATTCATGATTTCAAAATATTAATGATGATATGAAGAATTAATGAATTAATTAATTAATTA 957
QY 550 ttgaattctactcgtacctgttcgaaggcccaacgtaacgtaagctttggccaatgact 609
DB 958 TTAAATTTTATTTATTTATTTATTTAAGACAAATTTAAAGATGATGCTCATGATTTAT 1017
QY 610 tgcataatccatcattcaatttgaagatcgaagcgaagcttgagcattgagaagcttg 669
DB 1018 TGTCAATATCCTTTCAATCTTTAAATTCGTGCAATATGAATTTAGACCTACTTTAAAAACTT 1077
QY 670 gtctcgaatcgcgaagcctcgcgaacatcaagaagcattgtgggaagaatggaaga 729
DB 1078 GTGTTGGAATATGAAGAACCATTAGACAAATATTAAGATTAATGAGAAATATGGAAGAT 1137
QY 730 tatattaagaagaataagaagaccatcgaagaacattacgagctgatacgaagaatccaa 789
DB 1138 TACATTTAAAAAATATTAAGAACCATTAGAAATATTAAGATTAATTTAGAGAAATAG 1197
QY 790 aagaccatagacaagaataagaatgcgaaccgaaggaagaagaagaagttgtgaccag 849
DB 1198 AAACATATTGATTAATTAATTAAGATGCACTTAAGAAAGAAAGAAAGAAAGAAATATACCA 1257
QY 850 gcccaagtaagacctgtccatctataacaacagcttgaaagcccaatcaatcctacagc 909
DB 1258 GCTCATATGATGATCTTCTATTTATTAATTAAGATTAAGAAAGCACTAATTAATTAATTA 1317
QY 910 gtaactgagaagcgcataagacacctcaagaagaatgaaatatacaagaactgtcgcag 969
DB 1318 GTTTTGAAGAAACGTATTTACACTTTAAAAAATAAGAAACATTAAGAAATTAATTAATTA 1377
QY 970 aagattaagaatlaagaatctcctcgcgaagcgaagccttgaggaaacccctcaagcgtg 1029
DB 1378 AAGATTAATGAATTTAAAAATCCCGCCGCAATTTGGAATATACCAATTAATTAATTA 1437
QY 1030 ctggaacaagaacaagaagataagaagcagaagaagaagatcaagaagatccgcaaac 1089
DB 1438 CTTGATTAAGAACAAAAATTCGAGGAAACAGAAAGAAATTAAGAAATTCGCAAAAT 1497
QY 1090 attaatctacaatagattcctccttactgatacccttgagagctgagtaactgtgaga 1149
DB 1498 ATTAATTTTAATTAATGATGATTTATTTACTGATCCACTTGAATTTAAGAAATTAATTA 1557
QY 1150 gagaagaatlaagaatlaagacatctccgcgaagcgaagcgaagaagaagatcaacggaacct 1209
DB 1558 GAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1617
QY 1210 aatgaatcccaatgltgacgtlaccctcgtcttataaagatataccaacagcctc 1269
DB 1618 AATGAATATCCAAATGAGATTAATCTTCTTTGTCATTAAGATTAATTAATTAATTAATTA 1677
QY 1270 aacgagctcaatagctcgtggtgactgattaaccccttgatataagaagaacccctc 1329
DB 1678 AATGAATTAATTAATTTGCTTGTGATTAATTAATTAATTAATTAATTAATTAATTAATTA 1737
QY 1330 aagaatatacgaagacatlaagaagaagaagttatatacaagaagaatcaagaagaagatc 1389
DB 1738 AAAACATATATATCTGATTAATGAAGAAAAAATTCATTAATGAAGAAATTAATGAAGAAATTT 1797
QY 1390 aaatltgagaagaagaagaatltgagagltgacaagaagaagltgaagaagcccgcaagaag 1449
DB 1798 AAAATGAAGAAAAAATTTGATCTGATTAATTAATTAATTAATTAATTAATTAATTAATTA 1857
QY 1450 ctaacgataatacgaagaagatgaaagaagcgtggaacgagatctatgataccaatc 1509
DB 1858 TTAATGAATTAATTAATTAATTAATGAAGAAATTAATTAATGAAGAAATTAATTAATTAATTA 1917

Oy	1510	aacaaataacatcgaccctcgagccaaacttcgcggagaaatgatactggaacacggtactccttaca	1569
Db	1918	AAATATATATAGTTTAACTTAATTTACGAAAAATGATGGTAAAGATATTCATATATAA	1977
Oy	1570	gtggagaacatgaaaccccaataacccttgatccctatggaatctcaagcatatcctt	1629
Db	1978	GTGAGAAACTTACACCCATATACCTTTTGCCATCTTAAGAAAATTCATACATATATTT	2037
Oy	1630	gagaaactcacaagctctctaagiatatgagagactatctctgcgacaactctgtc	1689
Db	2038	GAAGAATTAAACAAGCTCTTAATATATATGGAAGATATATCTTTAAAGAAATATACAT	2097
Oy	1650	gagaaagaaactaaagatlatcacaagaactcctaagtaagtaagtcgaacagagatcgagac	1749
Db	2098	GAAGAAAGAAATTAATTAATTTATTAATAAATTTAAAGCAAAATGAAAAACAGATTGAACA	2157
Oy	1750	ctcttgtaagacttaagaaagatacggaagaacgttgttgtaagaagaagatcacaagaac	1809
Db	2158	TTAGTGTAAATATTTAAATAAGATGAGAACGCTTTTGAAAAAAAATTTACATTAACAC	2217
Oy	1810	gaaaataaacagatctgagaagaccccgagagctcccgatattgtctaaagtcgaagtcag	1859
Db	2218	GAATATTAACACAGATGAAAAAATTTTAAAGATATCGACATTTGTAAAGTACAAAGTTCAA	2277
Oy	1870	aaggtgctctcatgacaagaattgatctgaactcaagaagactcaactcaatctcgaagac	1929
Db	2278	AAATTTTATTAATTAATTAACAAATTTTGACGATTTAAATAAACCTCAATTTTAAATAAT	2337
Oy	1930	gtggaagttaaacaataatatacatctgycggaaatagatgaagcagagaaataagagaa	1989
Db	2338	GTACAAATTAACACTAATATATACATGTTCCCAATCTTCAACAAACGAATAATGACAA	2397
Oy	1990	ccatactaccctcatcgtcattccaagaagaagataagacaactgaagatctcatgccc	2049
Db	2398	CCTTATATTTAATTAATGCTTGAAAAAAGAAATGTATTAATTAAGTGTCATGCTACAG	2457
Oy	2050	gtcgaagcctgatcacaagaagaagaagaacatataaactcgaagacagtcagataac	2109
Db	2458	GTAAATATCATATGATTAAATGAGAAAAAATAACATTAACAGAAAGTCAATCGGATAT	2517
Oy	2110	tcgaagccttcacaagaagaagaagataaacccgaagcctcacacaagaaccccgagaacag	2159
Db	2518	TCGGAAACCATACMACGAAGAGAAATTAACAGGACAGCAACATACAAATCTGGACACAA	2577
Oy	2170	gcgcggtctcagctctcgaaagcgcatagcgtgcgaagctcgaagccaagaagcagaagcaga	2229
Db	2578	GCAGATCTGTTTAAAGAGGATGATCTCATACAAAGCACAAGCACAAGAACAAACAAACCA	2637
Oy	2230	caagctccagatgcccagatgccgtctccagagagctaaagctcaagatgctcacaaccacagct	2289
Db	2638	CACACACAGACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC	2697
Oy	2290	ccttgtaataagaagaccgggaatgfcagcgaactgagacttactctggaagcctcatag	2349
Db	2698	CCAGTAATATTAATTAATGAAAAATGTTTCCAAATTAAGATTATCTTGAAAAATTAATGA	2757
Oy	2350	ttccctgatacatcctacatctgcacacaatatatcctctgctctccacagcaactgac	2409
Db	2758	TTTTTAATATCTCATATATATATGTACAAATATATTTTGTTTCACTCAACATATGAC	2817
Oy	2410	gagaagatctcttaaacagtaacaagataaccaagaagaagagagataaactgctcctctgt	2469
Db	2818	GAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2877
Oy	2470	gataccatgagaccgctgctcaatctcgaagaacacatcccgatgatattctatgctc	2529
Db	2878	GATCATATTAAGCTTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2937
Oy	2530	gatacctcacaactctcctctcacaagtcatactgagataatcagaagaagatgctc	2589
Db	2938	GATAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2997
Oy	2590	tgcacaactgataaacctcaagaacacgacaagatlaagaacactctgaggaagctcag	2649

Db	2998	TTGTAATTTATATTAACCTTAAGGATTAATGACAAAATTTAAAAATTTATTTAGAGCAACGAAA	3057
OY	2650	aaggtctccaccctctgtttaaactctctctccagcttcacacacacgtctctctca	2709
Db	3058	AAAGTATCCACATCTGTAAAAACTCTTTCAAGTTCAATCAATGCACACATTATCATTAAACA	3117
OY	2710	cctcaagaacagcccggaagtcgagcgtctaaagacagacacctctcactcgacacaccta	2769
Db	3118	CCTCAGGATTAACCCGAAGTAAGGCAAAATGATGATACATCTTATCAAAATTTGAAT	3177
OY	2770	aactcactgaacatcgtcttggaacatccctgctctctcgcaagaataagaacatctacaa	2829
Db	3178	AATAGTTTAAATTTATTTTGGAAAAATATTCAGCTCTGGAAAAAACAAAATTTATACCAA	3237
OY	2830	gaacttatgtgacgaataacgctccggaactctctaaagaagaatctgaagaacgagac	2889
Db	3238	GNATTAATAGTCTCAAAAAAGTACTGAAAACCTTTATGAAAAGAGATTTTAAAGATATGTGAT	3297
OY	2890	acattctctaaacgagagcttcacttaactctgttgaatctctaaagccgatatcaactct	2949
Db	3298	ACATTTTATATGAACTCTTTTACAAATTTTGTAAATCTAAAGCTGATGATTTATATCA	3357
OY	2950	cttaacgatgaatctaaacgttaagaagcttggaaaggacatcaataagcttgaagaaga	3009
Db	3358	TTGAATGTGTGAATCAAAAAGGAAGAAATTGAAAGAAATTTAAATAATTAATAAAAACT	3417
OY	3010	ctgcgaacgagcttcgcgcctgtcaacaaagtaacaaactgaacttgagagactctgac	3069
Db	3418	TTACAGTATATCTATTGATTTTATATTAATTAATTAATTAATTAATTAATTAATTAATTCAT	3477
OY	3070	aagaagaagacagtcggaacaagtaaaagtcagatacaagaagtlgactctgctcaaggag	3129
Db	3478	AAAAAGAAACAGTTGGTAAATATATAAATGCATAATTAATAAACTTACTTTATTAAGAA	3537
OY	3130	cagcttgaagcaaaaccccaactactcgaacaatccgaacacgtaactcgagaactctca	3189
Db	3538	CAATTTAGATCAAAAATTTGAATTCCTTTAATAACCCAAACATGATTTTCAAAAATCTTTCT	3597
OY	3190	gtgtctctcaacgaagaaggaagccggaatctgcggaagacagagaacacactctgagag	3249
Db	3598	GTTTTCTTTTAAACAAAAAAGAAAGCTGGAATATAGCAAACTGAAAACACATTAGAAAAC	3657
OY	3250	accaagatctctctcaaacactcaaaagcctcgtcaagtattataatagcgagctctc	3309
Db	3658	ACAAAAATATTTGAAACAATTTATAAGGCTGTTAATATTAATATATATGATGATCACT	3717
OY	3310	cctctgaagactctctccgagagagacatccagaccgagataactaagccagctcgag	3369
Db	3718	CCATTTAAAAACTTTTAACTGAGAAATCAATCTCAAAACGAAAGATAATTAATCCAGTTTAA	3777
OY	3370	aacttcaaggtccgtctctaaagctcgaaggagcctgaagaagacaaacttgaacttgaagag	3429
Db	3778	AACTTTAAAGATATTAAGTAAATTAAGAAAGGAAATTAAGAGATTAATTAATTTACAAAAG	3837
OY	3430	aagaagctcagactcaaccctctaaagcagactcagacacttgaactcgcgagactcaagaagtc	3489
Db	3838	AAAAAATATATCATCTTATATCAAGGATTAACATCATTTAATTTGCGAATTTAAAGAACTA	3897
OY	3490	atlaagaacaaagactacacccggaatagcccaagcgagaataataiaacagacgttgaataac	3549
Db	3898	ATAAAAAATATAAAATTTATACAGCTAATTTCTCCAAAGTGAAGAAATAATACGATTTAACAA	3957
OY	3550	gcacgagatcttaacagaagttccgtccgtcggaggaagacagctgctccacgttggtgct	3609
Db	3958	GCATTAGATTTTCAAAAAATTTCTCCAGAAAGGAACAGATGTTGCAACAGTTTAACT	4017
OY	3610	gaatctgctccgacacacagagagcagcttcaaacctaaagaagcctgcatctcactcatgct	3669
Db	4018	GAAGTGTATCCGACATATTAGAACAAAGTCACAAAAGAACACAGCATCAACTATGTATA	4077
OY	3670	ggagcgaggtccaatattaccacatctcgaagcgtcgacagtgagtgctgatagctc	3729

D	4078	GGACGACAGTCTAACACAAATTAACACACTCACAATAGTCGATGATGACAGTACGTAACGTA	413738
Q	3730	atcatctgtccctatctcttcggcgagagcgagagagacctaagatgacctcggccaggtgtgc	3788
D	4138	ATCATAGTACCTATATTGGAGATCCGAAGAAGATTATGATGATTAGCACAACTAGTA	4197
Q	3730	acgggtgaggtctgcactctcttcggtgatttgataacattctgtccaaatcgagaagaa	3849
D	4198	ACAGAGAGAACGAGTACCTCCCTCCGTAATGATACATCTCTTCCTTCAAAATTTGAAAAATGAA	4257
Q	3850	taagaaatgtcctctctgaaacctctgcagagcgcttatagttctctcgaagaacagctg	3909
D	4258	TATGAGGTTTATATTAAAACTTTAGCAGGTCGTTTATAGAACTTTAAAAAACAAATTA	4317
Q	3910	gagataacgctgataccctccaatgtctcaacgtaagaagacattctgaacacgcgccttatt	3968
D	4318	GAAATTAACGTTATGACATTTAATGTTAATAGTAAGGATATTTTTAATTCACGATTTAAT	4377
Q	3970	aagagagaaaattccaagaacgtcttgagagcgacttgattccctataaagacgtgacc	4029
D	4378	AAAGGTGAAAATTTCAAAAATGTTTTAGATTCAGATTTAATTCATATTAAGATTAACTTAACA	4433
Q	4030	tcacctactacgtgtgcgaagaccatacaagttccctcaataaagaagaagggataaa	4083
D	4438	TCAAAGTATTAATGTGTGCAAAAGATCCATATAAAATTTCTTAATAAGAAAAAGACATATA	4497
Q	4090	ttctgtctagttcaactatatacaagagctccatcgacacacgataccaattctgctat	4149
D	4498	TTCTTAAGCACTTATATTATTTAATAGATTCATAGATACGATATATAATTTTCCAAAT	4557
Q	4150	gattgtcgtgggtattacaagatccctgagcgaaaaatacaagltctgaccttgactcat	4209
D	4558	GATGTCTTGATATATTAATAAATTTATCCGAAAAATTAATTCAGATTAGATTCAAAT	4617
Q	4210	aaaaagatatacaacgctgaagcgagcgagaatgaaaaaatactgcacctctctgataac	4268
D	4618	AAAAAATATATCAACGCAACAAACAGGTAAGAAAAATGAACAAATACCTCTCTTTTAAACAT	4677
Q	4270	atcgaaacccgttacaagacagctgaaacgaacaaatcgacctctctgataccaacctgag	4329
D	4678	ATTGAGACCTTATATTAACAGATTATAGTAATGAATTAATGATTATTGTTGTAATTCATTAGAA	4737
Q	4330	gccaaagtccctcaactatacttaacgagaagagcaatgtggaagttaaatacaagaagctg	4389
D	4738	GCAAAATGTTCAATATATACATATGAGAAATCAAAAGTGAAGTTAAAAATAAAGAACTT	4797
Q	4390	aacctactcaaaaacaatccaagacaagctgagcaattccaagaaaaataacaattctgc	4439
D	4798	AATTACTTAAAAACAATTCAGACAAATTTGGCAGATTTTAAAAAAAATPAACAATTTGCTT	4857
Q	4450	gaaattcgaaacccgtcttaacgatttaaacacaacaactctctgaacaaagtcttctcc	4509
D	4858	GGAAATCTCTGATTATTAACAAGATTATTAACCAATTAATCTTATGCAAAAGTTCCTTAGT	4917
Q	4510	actgtgacatggttctcgaaaaacctgcaccaaacaggtgtcgaagcaatctgcgcgaac	4569
D	4918	ACAGGTATGTTTTGCAAAATCTTGCTAAAAACCGTTTTATCAATTTACTTGATGGAAC	4977
Q	4570	ctggaaggaatgtctgaaacatctcccgacaacaaatgtgtaagaaaaaacgtgccccgaat	4629
D	4978	TTGGAAGGTATGTTAAACATTTTACCAACCAATGCGTAAAAAAAACAAATGTCCACAAAT	5037
Q	4630	agcgccgttctcaagcaatctgcagcgcgagaggtgaaggtgtctctctgaactacaa	4689
D	5038	TCGTGATGTTTCAACACTTTAGATGGAAGAAGAAATGTAAATGTTTATTAATTATTCAAA	5097
Q	4690	caagaagagataaggtgcgtgcgaacccaacaccttaccctgcaatgaaacacatgtgcgg	4749
D	5098	CAMGAAGTATTAATGTGTTGAAAATCCAAATCCTACTTGTAAAGAAAATATATGTGGA	5157
Q	4750	tgtgaacgcatgtctaaatgcacccggaggaagacagcgcttcaacggaaagaataacaca	4809
D	5158	TGTTATCTCAAGTGCCAATATCTACCGAAGAAGATTACAGTACGACGGAAGAAAGAAATTCACA	5217

QY	4810	tgcgagtgctactaaagcccgaccctccatccacctcttcgaaggagatctttgtccaccgctc	4869
Db	5218	tgTGAATGACTAAACCGCATTCCTTATCCACCTTTTCGAGTGATTTTTCGACAGTTCCTC	5277
QY	4870	aattccctggagatccctccctcgtcgtacccaccatgcagtcgataccctgaaagatcatc	4929
Db	5278	AACTCTTAGGAATTCATCTTATTATTAACTACATGTTAATATATTATACAGTTTCATTAA	5337
QY	4930	tagatcgatg	4939
Db	5338	AAATGTAGG	5347
RESULT	4		
PGP195A			
LOCUS			
DEFINITION	PGP195A 7038 bp mRNA INV 03-APR-1995		
ACCESSION			
VERSION	X15063.1	GI:9896	
KEYWORDS	antigen; glycoprotein; gp195 gene; surface antigen.		
SOURCE	malaria parasite P. falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.		
AUTHORS	1 (bases 1 to 7038)		
TITLE	Myler, P.J.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (14-APR-1989) Myler P.J., Seattle Biomedical Research		
AUTHORS	Institute, 4 Nickerson Street, Seattle WA 98109-1651, U S A		
TITLE	2 (bases 1 to 7038)		
JOURNAL	Myler P.J.		
REFERENCE	Nucleotide and deduced amino acid sequence of the gp195 (MSA-1)		
AUTHORS	gene from Plasmodium falciparum Palo Alto PLF-3/Bl1		
TITLE	Nucleic Acids Res. 17 (13), 5401 (1989)		
JOURNAL	89345116		
MEDLINE			
FEATURES			
source	Location/Qualifiers		
	1..7038		
	/organism="Plasmodium falciparum"		
	/strain="Palo Alto PLF-3/Bl1"		
	/db_xref="taxon:5833"		
	/dev_stage="trophozoite/schizont."		
	1592..1648		
	/product="signal peptide (AA -19 to -1)"		
	1593..6511		
	/codon_start=1		
	/product="gp195 surface antigen preprotein"		
	/protein_id="CAA33163.1"		
	/db_xref="gi:9897"		
	/translation="MKIIIFLCSPLFIINOCYTHESYOELVKLEAEVAVLTGYS		
	LFOEKAVLNGCTGTAVTTPSGSKGSAVSGSGSAVSGSAVSGSAVSGSAVSG		
	GGSGNSKRTNPSDSSDSDAKSYADLKHRYVNTLLITKELYPOLFDTLNMILCLD		
	ILHGKYLIDGEEIENELLYKLNFFDLIRAKLVNVCADYCOIPFNKIRANEIDVLA		
	KLIVGKYLDPDNIKDNVGMEDYIKKKKTITENINLEIESKTIKKNKATKEEEKK		
	KLYAOQDLSTYNRQLEEAHNLISVLEKIDITLKNNEIKELDLINKNIPANS		
	NTPLDLDNKKIEHEHEKEIKELKTIFNDISLFTPLDELEYLEIREKNKIDISAKK		
	ETKSTEPNENPVGVTPY2IYNDINMLNENSGDILINPDYIKRPSNITYOTNER		
	KFINTEKEIKITEKKKIESDKSIEDSKSLNDITKEKILNLEIYDSKPNNDITR		
	FEKMGKRISTKVEKLTHNTPFASYENSKHLEKLYLAKYMEYSLNIVYERELK		
	YKNLISKIENIEIETLVENIKKDEQOLEEKKTKTPOENKPFDEILVESDVKVQV		
	MNKIDELKLTQILIKNLEKIHNLIVNPSKYOENOEYPLVILVKEIDKILKMPKVE		
	SLIIEKKNKITEGOSNPESTGETIGLQATITPGQOAGSALGDSVOAQAOKOAK		
	OPVPEVPEKKAOVPTPPAPVNNKTNVSKYLDLEKLYEFILNSTYCHKYLISHS		
	MNEKTLQYKIKTREESKISCDPDLILFNQNNITPWYSNFDLSNLSQDIMEIYH		
	KEMGNCILKLDNDKIRNLLEAKKVSATVTLSSSQPLSLTPQDPPEVSANDDTIS		
	HSTNLNSLKLFEINILGKNKNTYQELIGQSENEFEKILKQSDIFYNKSPFNFW		
	SKADILSLNDESKRKLLEEDINKLQTLQSPDLNFKYKLLEKLPDKKATVGRYK		
	QIKLUTLILKOLESKLNSLNPKRYVLNLFQSVFKKKEAEIETENLENTKYLILK		
	KGLVYVWESSPLKTLSEESIOEDNWNVSLNENLVLSKLEGLKDNMLEKKLISV		
	SSGLHHILAELEKRYIKKNKNTYGNPSNNDVNNALSKYKKEFDPGIDVATVSESG		
	DTLEOSQPKRPASTHVGABESNTTTSQWVDEVDVILVPIFGSEEDYDGLQGVV		
	EAVPYSVINDILSKIENEYEVLYLKLPLAGVYRSKLQLENNVWFNVNVDILNSR		
	KEKRFKKNLSDLLPYKDLTSSNVVDYFLKREKDKRLSSYNIKQSIDINDIN		

QY	1990	ccatactaccctcatcgttaccgaagaagatagacaacctgaaagtgctatgccaa	2049
Db	3572	cccttttttttttaatttggttgtaaaaaaaatttgatataattttaaagtcttcgtccatg	3631
QY	2050	gtcgaagccgtgacccaacgagagaagaagaacatttaaacctggaagcgatgcacatac	2109
Db	3652	gtagaaatcattgataaattgaaagaaaaaaaacattaaaaacgaaaggtcatcgatnat	3651
QY	2110	tcgaagcctccacagaagagagataaccggaagctaccaccaagcccgaacaaag	2169
Db	3692	tgcgaaccattcacccgaaggaataatpacagcagcaactatcAAAAAACCtgcgaacaa	3751
QY	2170	gcgcgttcagcttcggaagcgatagcgttggaagcctcaagcaagaagcagaagcagca	2229
Db	3752	gcagagatcgtcctttagaaagagattctcagttcaacgcaaaacgcaaaacaaacaa	3811
QY	2230	cagcctccagtgccagtgcccggttccagaagctcgaagctcaagtgctcacaaccagc	2289
Db	3812	caacccacagatgacagatgacacagatgacgaagaagaaagcacaagctccaaacac	3871
QY	2290	ccctgtgaataacaaagaccggaatgctcagcaaacctgagctactcttgaaagctatag	2349
Db	3872	ccagtaataataataaactgtgaaaaatggtttccaaatttgatattcttgaaaaaattatg	3931
QY	2350	ttccgtgatacatcctacatctgcacaatatatccctgcttcacagacatgaac	2409
Db	3932	tttttttaattcttcattatattatctcmaattatatttggtttcacactcactcactg	3991
QY	2410	gagaagattccttaacagatacaagataacaaaggaagagagataaacctgctcctgt	2469
Db	3992	gaaaaagatttttaaaacaaatttttaaaatttacaagaagaaagaaacgaattttgaatg	4051
QY	2470	gatacaactggacgcgcgtgttcaatcaccagaacaacattccggtatgatctatctc	2529
Db	4052	gatacatttagacttattgtttattatracaaatracatcctgtaattgattttctatgtt	4111
QY	2530	gatagcctcaacaactctctctctcctaactgttcatatgagatalatgaaagagatg	2589
Db	4112	gattgacttttaaaacattactttttacacacttttatgaaaattttatgaaaaaaatg	4171
QY	2590	tgaacactgtataaaccaagaacagacagaatlaagaacactcttgagagagctcaag	2649
Db	4172	tgtatattttatrtaaactttaaggaattatgcAAAAATTTTAAATTTTAAAGAAAGCG	4231
QY	2650	aaggtctccacctctgtlaaacctctctctctcagctccatgcaacacactgtctcca	2709
Db	4232	aaatttatccacactctgttaaaaaactctttcaagttcatcattgcaacactttatcca	4291
QY	2710	ccataagaacagccggaagtgagcgtlaacgacgaacactctcacttgaccaactta	2769
Db	4292	ccctagaatatacccggaagatgtaaaatgataatgatactacattcttcaaaatttgat	4351
QY	2770	aaccactgaaacactgtttggaacacccgtctctcgcgaagaataaagaacattacca	2829
Db	4352	aattgatttaaaattatgtgaaacattatgagcttggaaaaaaacaaataatattacca	4411
QY	2830	gaacttatggacaagaacatcgctccggaacctctacgaaagatactgaaagacgac	2889
Db	4412	gaattatattagctcaaaaactactgtaaaaactttttatgaaaagattttttaaagatg	4471
QY	2890	acattcctagaagagccttactactactcgtgaaatctaaagccgagatgataactc	2949
Db	4472	acattttttatattgaaactttttacaaatttttgtaaaattcttaagctgagattttatcca	4531
QY	2950	cttaagatgaatctaaacgtlaaagcctggaagagacatcaatlaagctgaaagaaga	3009
Db	4532	tttgaaattgataaatacaaaaagaaagaaattagaagaagattttatatttttaaaaact	4591
QY	3010	ctggaactgagcttgagccctgttcaacaagatacaaacctgaaactggagagactctgc	3069
Db	4592	tttACAGTTATCAATTTGATTTATATATATTAATTAATTAATTAATTAATTAATTAAT	4651

OY	3070	aagaagaagacagtcgcggaagtataagaatgcagatacaagaagtgcctctgcctcaaggag	31289
Db	4652	AAAAAAGAAACAGTTGGTAAATATATAAATCCAAATTAAAAAACCTTTATTTAAAAAGAA	47111
OY	3130	cagctctgaagcaaacccaactcaactcctgaacaatccgaaacaactctacgtgaactctcca	31889
Db	4712	CAATTAGAAATCAAAATTTGAATTTCCTTTAATTAACCCAAAGCATGTATTACAAACCTTTTCT	47711
OY	3190	gtctctctcaacaagaagaagaaagccgagatcgcgcgagacagagacaacatctctgaagac	3249
Db	4772	GTTTCTTTAACAAAAAAGAAAGCTGGAATATGACGAATCGAATAACATTAACATTAGAAAC	4831
OY	3250	accaagatctctctccaaacactacaagagccctgcgaagtatatagaatggcgagctctc	3309
Db	4832	ACAAAAATTTATTGGAACAATTTATAAGGACTGTTAAATATTTAATATGTCGAATCATCT	4891
OY	3310	cctctgaagactctctccgcggagagagatccagccgagatacaactacgcgcgcgcctgcag	3369
Db	4892	CCATTAAAAACTTTTAAGTGAAGATTCATTTCAAACGAAAGATTAATTAATCCGTTTAAGA	4951
OY	3370	aacttcaagagccctgcctctaaagctcgaaggaagcctgaagacaacatctgaagcctgaagag	3429
Db	4952	AACTTAAAGATTAAAGTTAAATTAGAAGAAATTAAGAGATTAATTTAAATTTACAAAG	50111
OY	3430	aagaagctcagctacccctctcagcgagctgcataccctgcgcgcgcgcgcctcaaggagtc	3489
Db	5012	AAAAATATCATCTTATCAAGGATTCACATCATTTATTGTCGAATTTAAAGAAAGTA	50711
OY	3490	attgaagaacgaagactcacccgcgcaataagcccgagcgagaataatacagagcgtgaatcac	3549
Db	5072	ATAAAAATTAATAATTTTACAGTTAATCTCCAGAGGAAATAATACGGATTTTACAT	51311
OY	3550	gcactggaatctctcaagaagctctccctgcctcgaagagacagatctgcgcactctgctgct	3609
Db	5132	GCATTAGATCTTCACAAAAAATTTCTCCCGAAGGAAGACAGATGTTGCAACAGTTGTAAGT	5191
OY	3610	gaatctgcctcgcgacacacgcggagcagctcacaactaagaagcctgcatactcaatctgc	3669
Db	5192	GAAATGTGATTCGACACATTAGAACAAAGTCAACCAAAACACGACATCAACTCATCTGTA	5251
OY	3670	ggagccgagctccaatacaatctaccacactccgaagcgtcgaagatggaagtcgaatgagctc	3729
Db	5252	GGAGCAGAGCTTAACACATTAACACACATCCAAATAATGTCAGATGAAGTACATGACGTA	53111
OY	3730	atcatctgcctatctctcgcgcgaagcgcgaagagagactacgaatgaactcgcgcgaagtctgc	3789
Db	5312	ATCATAGACTATATTGGAGAAATCCGAAGAAGATTAATGATGATTTAGCAACAGTACTA	5371
OY	3790	accggtgaagctctcactcctctccgtgattgtataacatctctccaaatccgagaagca	3849
Db	5372	ACAGAGAAAGCAGTAACTCTCTCCGTAATTTGAATACATCTTCTTAATAATTGAAATGAA	5431
OY	3850	taagaagtgactctatctgaaacctctcggaagcggtctatagctctcttaagaagaagctg	3909
Db	5432	TATGAGTTTTATTAATTTAAAACTTTAGACAGGCTTTATAGAAATTTAAAAAACAAATTA	5491
OY	3910	gagaaatcgtgaactctcaatctgcaacgtcgaagacatctgaacagccgcttcaat	3969
Db	5492	GAAATTAACGTTATGACATTTAATGTTAATGTTAAAGATTTTAAATTCACGATTTAAT	5551
OY	3970	aagagagaanaattccaagaagcgtctcgagagcgagctgcatactccataaagacctgacc	4029
Db	5552	AAAGGTAAAAATTTCAAAATGTTTATAGAAATCAGATTTAATCCATTTAAAGATTTAA	5611
OY	4030	ctccctactacgtctgcgaagagcccatatacagttcctcataaagaagaagagatataa	4089
Db	5612	TCAAGTAAATTATGTGTCAAAAGATTCATTAATTTCTTAAATAAGAAAAAGAGATTA	5671
OY	4090	ttctctgctagttaacaactatcatcaagagactccatcgaacccgatatcaaatcttcgcatc	4149
Db	5672	TTCTTAAGCAGCTTTAATTAATTAATTAAGGATTCATTAATGATACGGATTAATAATTTGCAAT	5731
OY	4150	gagctgcgtgggattatacaagatccctgcgcgaagaatacaaatctgcgacctgcactatc	4209

[illegible]

Source	1. 4920	/organism="unidentified"	/db_xref="taxon:32644"
BASE COUNT	2194 a	589 c	679 g 1458 t
ORIGIN			
Query Match	57.0%;	Score 2814.4;	DB 9; Length 4920;
Best Local Similarity	73.3%;	Pred. No. 0;	
Matches 3604; Conservative	0;	Mismatches 1316;	Indels 0; Gaps 0;
10 atgaaatcattcttctctctgtcttcattcgtttttatcatcaatactcagtcgctg	69		
1 atgaaagatcatatpctttttatgttcattctctttttttatataaataacacaagtgtga	60		
70 accccagcaatccatcaggaagcttgatlaaagaaacgtggaagcttggaaatgcctcctt	129		
61 acacatgaaagatattacaaagaaactgtgcacaaaaaactagaagcctttgaaacatgcagatttg	120		
130 accggaatacagcctgtlccagaagagaaagatgctgtcgtatgaatgaagggcagctgagcag	189		
121 acaggttatagcttttttttcaaaaagcaaaaatggttatataaagaagaaacagtgacma	180		
190 gccgttacacacagcaacacccggttctaaaggtctgtgtcagcgggtgtcccggttgg	249		
181 gctgtttacacagtagaacaacctgttcamaaggggttcagtttcagctggttcagctggcc	240		
250 tctgtggcctctgggggttcgcgtgcgcctccggccggcagcgctggcatcaggtgctcagtg	309		
241 tcagtttgcttcacgggtgcctcagttgcttcagctgcttcagctgcttcagctgcttcagtt	300		
310 gcaagcgcggtgtccgggagacagtcgaagaacacatccatctgacacactcagcagttcc	369		
301 gcttcacagtggttcaggtatattcaagagctacaaatccttcagatattcaagtgatcca	360		
370 gagcccaagctcctacgcgcagcctcaagcagcgatggaactatctcctcatcaatcaag	429		
361 gatgcataattcttacgcgtattttaaacaacagtagcgaataattactgtttactatcaaaa	420		
430 gagctgaagctaccacagctgtgtcgaacctcactatcatatgctgacactgtgtgaatc	489		
421 gaactcaaatatccttcacactcttctgttgaatttaactaatcatatgttactctgtgtgataat	480		
490 atccatgctcacaatcctatctgaatctgacggtctcgaagagatcaatgaactccgtctcaag	549		
481 attcatatggtttcaaatattttatttgatgcatatgaaagaattatgaattattattatataa	540		
550 ttgaatttctactcgaactgtcgaaggccaaactgaaatgacglttgccaaatgcatat	609		
541 tttaactttttatttgattttatttaagacacaaatatttaaatgattatgctcattatgattat	600		
610 tctcaaatccatcaatatttgaagatcgaagccaaagctggagctcttgaagaagtgtg	669		
601 tctcaaatccatcttcaatttttaaatctgtgcaaatgaattatgacctacttttaaaaacttt	660		
670 gctctcgaatctcgaagcctctcgcacaacatcaaggaacaatgttggaaagaatctgaagaat	729		
661 gcttccggttatgaaacacacattatgacatattttaaagatattgtgaaagaaatgcaagatt	720		
730 tatataaagaataagaacacatcgaagacatlaacgagctgtagcgaagaatccaaa	789		
721 tacaattttaaataaaatvtaaaaacacattagcaaaatttttaaatgaattttatgaaacaaactag	780		
790 aagacacatagaacaaataagaatgcaacccaaggggaagaaagaagaagtctgtacag	849		
781 aaacacattgattttaaataatgaaatgcacctttaaagaagaaagaaataattatataccaa	840		
850 gcccgctacgacctgtccatctatacaacaacagcttgaagaagcccatataactcagc	909		
841 gctcatatagatcttctttttacataataacatttagaagaagcacaattatttaattatgaac	900		
910 gtacttggaaagcgcatagacaacctcaagaagaatgaataatcaagaagactgtctcgac	969		

Db 901 GTTTTAGAAAAAGTATGACACTTTAAAAAATGAACATTAAGCAATTACTTGAT 960
Qy 970 aagattaatgaatlaagaatctccgcagcaactctgggaagcccttaacagcgtg 1029
Db 961 AAGATTAATGAATTAATTAATCCCGACGCGCAATTCGGAATACACCAAAATCTCTC 1020
Qy 1030 ctgcaagaagaagaagatagagagcagcagaagaagatcaagaagatcgccaacg 1089
Db 1021 CTTGATTAAGACAAAAAATTCGAGGAGACAGAAAAAGAAATTAAGAAATTCGCAAACT 1080
Qy 1090 attaatlaacaatagatctctcttactatgaccttgagctgagtaacttctgaga 1149
Db 1081 ATTAAATTAATTAATTAATGATTATTAATTAATTAATTAATTAATTAATTAATTA 1140
Qy 1150 gagaagaataagaatlaagacatcccgcaagctgagcagaagaagatcaacagcact 1209
Db 1141 GAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1200
Qy 1210 aatgaatcccaatgltgagcgtacccctctgcttataacgatalcaacaacgtctc 1269
Db 1201 AATGATATCAAAATGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1260
Qy 1270 aacgagctcaatagctctggtgactgacttaacccctcgatatacgaagaacccctc 1329
Db 1261 AATGAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1320
Qy 1330 aagaatctacacagacagatgagaagaagaagttatcaacgaatacgaagaagacgc 1389
Db 1321 AAAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1380
Qy 1390 aaatctgagaagaagaatctgagagtgacaagaagaagtaacgaagccgcaagaagt 1449
Db 1381 AAAATGAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1440
Qy 1450 ctaaaagatatacctaagaagatgaaagcgtcgtgaacagagatctatgatlccaatlc 1509
Db 1441 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1500
Qy 1510 aacaataacatcgaccctgacacactcgcgaagaatgaggaagaacgtaactctcaaa 1569
Db 1501 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1560
Qy 1570 gtgaggaacctgacacacataatcttgcactctatgagaatcttaagaataactt 1629
Db 1561 GTTGGAACCTTACACCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1620
Qy 1630 gagaagctacacaagctcttaagatatacgagagactatctcctggggaacatgtgtg 1689
Db 1621 GAAAGTTTAACAAAAGCTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 1680
Qy 1690 gagaagaacataagfatatacagaatctcatalaagtaagatcgaaagacagatcgagc 1749
Db 1681 GAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1740
Qy 1750 ctgttgagaacataagaagaatgagaagacagltgttgagaagaagaatctcaaaagc 1809
Db 1741 TTAGTTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1800
Qy 1810 gaaataaacccagatgagaagaatcccgaggtctccgatatgttlaaagtaagtgag 1869
Db 1801 GAAAAAATTAACAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1860
Qy 1870 aaggtgctctcatgacacagatctgaagcaagaagaagacataactctcttgagaagc 1929
Db 1861 AAGATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1920
Qy 1930 ctgaggttataaacaataataatagatgagcgaagatgataaagagagaataagcagaga 1989
Db 1921 GTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1980
Qy 1990 ccatctactcatcgtactcaagaagaagaatagacaaactgaagtgctcatgcccaga 2049
Db 1981 CCTATTATTATTATGTTGTAAGAAAGAAATTAATTAATTAATTAATTAATTAATTAAT 2040

Qy 2050 gtccgagagccctgatacaacgagaagaagaacattaaactgaagagacgtcacataac 2109
Db 2041 GTTGAATTCATTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2100
Qy 2110 tccgagccttccacagaagaagatataacccgagacaggtctacacacgaagcccggaacag 2169
Db 2101 TCGGAATTCATCAACCGAGAGAAATTAACAGGACAAGCAATTAACAAACCTGAGACAA 2160
Qy 2170 gccggttcaagctctgagaagcgatagcgtgcaagctcaagcaagaagcagaagcgca 2229
Db 2161 GAGGATTCGCTTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2220
Qy 2230 cagccctcagtgccagtgccgcttccagagagctaaagctcaagtgctcacacacagct 2289
Db 2221 CAACCCACGATACAGTACCAATGATTAATTAATTAATTAATTAATTAATTAATTAAT 2280
Qy 2290 cctgtgaataaagaagaagcagagatgcaagcaactggaactcttgagaagctctatgag 2349
Db 2281 CCAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2340
Qy 2350 ttcctgataatatacctacatctgcccacaatatatcctcgtctcacaagcactatgaac 2409
Db 2341 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2400
Qy 2410 gagaagatcttaaaacagtaacagataaacgaagaagaagatgaactgctcctgtg 2469
Db 2401 GAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2460
Qy 2470 gatcaactggaacctgctgttcaatatacagaacaacatcccgatgatatgatatgctc 2529
Db 2461 GATCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2520
Qy 2530 gataagcttaacaatctctcctcaacgtctcaatgagatatagaagaagatggtc 2589
Db 2521 GATAGCTTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2580
Qy 2590 tgcgaacctgtataaactcaagaagaagaacgaagaatgaagaacactcttgagaagaagc 2649
Db 2581 TGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2640
Qy 2650 aaggtctcaacctgtgtttaaactctctctccagctcctcaagcaacactgctctcaca 2709
Db 2641 AAGATTCACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2700
Qy 2710 cctcaagaagaagccggaagtgagcgtcaagcagacactcgaactcgagccacttaac 2769
Db 2701 CCTCAGATTAACCCGAAAGTAAAGTGAATGATGATGATGATGATGATGATGATGAT 2760
Qy 2770 aactcactgaacactgttgagaacatccgtctcctcgcgaagaataagaacataccaa 2829
Db 2761 AATAGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2820
Qy 2830 gaacttatggaacgaagaatcgccggaactctcagaagaatctgaagaagcagcagc 2889
Db 2821 GAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2880
Qy 2890 acattctaaagagagactctcaactctcgtgaatctcaaaagccgagatgataactct 2949
Db 2881 ACATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2940
Qy 2950 cttaaagatgaatctaaagcgtgaagaagctggaagaagacataaagctgaagaagaca 3009
Db 2941 TTGAATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3000
Qy 3010 ctgcaactgagctcgaccctgtaacaagaatcaaaactggaagactctctcgac 3069
Db 3001 TTACAGTTTATTCATTTGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 3060
Qy 3070 aagaagaagaagctcggaagatataagatgacatgaagaagtgaactgctcaagag 3129
Db 3061 AAAAAGAAAAAGCTTGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3120

Oy	3130	cagcttgaagaacaaacccaactcgaactcgaaacaatccgaaacacgtctctgcgaactccca	3189
Dd	3121	CAATTAGAACATCAAAATGGAATTTACTTAATAAACCACAAACGATGTATTCACAAAACCTTTTTT	3180
Oy	3190	gtctcttcaccaagaagaagaagagccggatctgcgcgacagagaaacactctgtgaagac	3249
Dd	3181	GTTTTCTTTTAACAACAAAAAAGAAGCTGGAATATGCAGAACTCGAAACACATTAGAANAAC	3240
Oy	3250	accaagaattctctccaacaacatacaaaaggccctgcctaaglatataatgygcgtctct	3309
Dd	3241	ACAAAAATATTTATGAAGAACATVTATAAGAGCCTTGTTAAATATTATTAATAGTGGAATCANCT	3300
Oy	3310	cctctgaagacctctctccggaggagcatccagaccgagataactacgccaagcctcgag	3369
Dd	3301	CCATTTAAAACTTTAACTGAGAGATCAATTTCAACAGAAAGATAAATTAATTAATTTGGAAGA	3360
Oy	3370	aactccaaggtctcgtctaaagctcgaagccaagctbaagagaacactgaacctgaagag	3429
Dd	3361	AACTTTAAAGTATTAACTAAATTAGNAGGAAATTTAAAGATAAATTTAAATTTGGAAGAAG	3420
Oy	3430	aagaagctcagctacccctctcgaagcagctgcacactgatccgcgagctcaagaagctc	3489
Dd	3421	AAAAAATATCATACATTATCAAGTGGATTACATCATTAAATTTGCGAATTTAAAGAACTA	3480
Oy	3490	altaagaacaagaactcacaccggacaalaccocaagcggagaataatacagacgttgataac	3549
Dd	3481	ATAAAAATATAAAAATTTATACAGCTAATTTCTCCAAGTGAAGAAATAAATACGATTTAACAT	3540
Oy	3550	gcactggaaactttacaagaaggtctccgctctgaaggagaagagatgtccacagtgtgct	3609
Dd	3541	GCATTTAGAAATCTTACAAAATAATTTCTCCAGAAAGGAACGATGTTGCCAACAGTTGTAAGT	3600
Oy	3610	gaactgtgcctcgacaacactcgygacagctctcaacctaaagaacctgcatactatctc	3669
Dd	3601	GAAGTGGATCCGCACATATTAGAACAAAGTCAACCAAGAAACACGATCAACTCATGTA	3660
Oy	3670	ggagccgagctccaatacatalaccacatctcgaagcgtgcagatgtagtgcgtatgcgtc	3729
Dd	3661	GGACACAGGCTTAACCAAAATMACATCACAAATATGTCAGATGAAGTATGATGACGCTA	3720
Oy	3730	atactgtgcctatctctccgcgcgagcggaggagactgaagatgcagctgcgcgaagtgctc	3789
Dd	3721	ATCTTAATACCTTATATTTTGGAGAAATCCGAAGANAGATTATGATGATTTTGGACAAAGTATGA	3780
Oy	3790	accggtgaaggctctcactcccctccgcgtgatgtglaaacattctctccaaaactcgagaacga	3849
Dd	3781	ACAGGAGACAGAGTAACCTCTCTCCGTAATTTGATTAACATCTTCTTAATAATTGAAANATGAA	3840
Oy	3850	tagaagltgcctacatctgaaacctctgcgaagcgtctabagtlctctcaagaanaacgctg	3909
Dd	3841	TATAGGCTTTAATTAATTAAMAACTTTAGCAGGCTGTTTATAGAACTTTAAAAAACAAATTA	3900
Oy	3910	gagaaatagctgagtcoccttaaatgtcaacgfygaaggeaatcttgaacagccogcttaac	3969
Dd	3901	GAAATATACGTTATATGACATTTAATGTTAATGTTAAGATATTTTAAATTCACGATTTAAT	3960
Oy	3970	aagagagaaatctcaagaagcgtctcggagagcgagctgatctccctcaaaagacctgcgc	4029
Dd	3961	AAAGGTGAAATTTTCAAAAATGTTTTAGATCAAGATTTAATTCATTAATAAGATTTTAACA	4020
Oy	4030	tctcttaactagctgtlcaagagccacatacaagttcctctcaataaagaagaagatatnaa	4089
Dd	4021	TCAAATATATATGTGTGCAAAAGATTCATTTAAATTTCTTAATAAAGAAAAAGAGATTA	4080
Oy	4090	ttctgtcgtaggttaagaactatacaagagctccatccgcagaccgataataactllcgctaac	4149
Dd	4081	TTCTTAAGCATATATATATTTATTAAGATTCOAATATGATACGGAATATTAATTTGCAAT	4140
Oy	4150	gatlgtctgggagtatlacaaagatcccyagcgaaaaaaeaaagctgcagccttgactcata	4209
Dd	4141	GATGTCTTGATATTAATAAATATTTATCCGAAAAATATTAATTCAGATTTAGATTTCAATT	4200
Oy	4210	aaaaagatatacagaataagcaagcgagaaibgaanaatabctgcctctccctgataaac	4269

[illegible]

QY	1630	gagaagctcaccaaagctctctaaagtatabgagagactatctctcgcggaacatgctgctg	1689
Db	1836	GAAGATTAACCAAAAGCCTTAATATATGCGAAGATTAATCTTTAAAGCAATATAGTACTT	1895
QY	1690	gagaaagacataaagattctaaagaatccataagtaagaatcgaagaagatcgagagc	1749
Db	1896	GAAGAAAGATTTAAATATTTTAAAAATTTTATATAGCAAAATATGAAATATGAGATTGAACA	1955
QY	1750	ctgtctgagaacatctaaagaagaatgaagacagctgtctggaagaagaattactaaagac	1809
Db	1956	TTAGCTGAAATATTTAAAAAAGATGAAAGAACAGCTTTTGAAAAAAAATTTATCTAAAGAC	2015
QY	1810	gaataataacagatgagaagatcccgagagctcccgatatgttaagtcgaagtcag	1869
Db	2016	GAATATTAACCCAGTGTGAAAAAATTTTATGAAAGTATCTGACATTGTGTAAGTCAAGTTCAA	2075
QY	1870	aaggctgcctccatgaaacaagatctgaactcaagaagaacccaactcatctgaagac	1929
Db	2076	AAAGTTTTATTAATGAAACAAATTTGACGAATTTAAAAAACATCCATGTGATTTTAAAAAT	2135
QY	1930	gtggaattaaaacaataataatactgtgcgcgaagttctaaagcagagaataagcagaa	1989
Db	2136	GTGAATATTAACACTAATATATACATGTTCCCAATTTCTTCAAAAAGAAAAATATAGCAAGAA	2195
QY	1990	ccatctaccctcatcgtaactcaagaagaagatagacaacaatgaaagcttctatcccaa	2049
Db	2196	CTTATTTATTTAATGTGTTGAAAAAAGAAATTTGATTAATTTAAAGTGTTCATGCTCTAG	2255
QY	2050	gtcgaagacgctgatacaacgaagaagaagaacatctaaactgaagaagctcagataac	2109
Db	2256	GTGAATATCATGTTAATATGTAAGAAAAAATAAATTAACCAAGCAAGTCAATCGATAT	2315
QY	2110	tcgaagctctccacagaagaagaagataaacgcgaacggtccaccacaagcccgagacaag	2169
Db	2316	TCGGAACCATTCMACCGAAGAGAAATATACAGACACACTACAAAACCTGGCAACAA	2375
QY	2170	gcgcggtctcaagctcgaagcgatagcgtgtcgaagctcgaagcacaagaagcagaagcagca	2229
Db	2376	GCAGAGATCTGCTTTAGAGAGATGTCATGATCAAGCAACAGCAAGCAAGAAACAAAGCA	2435
QY	2230	cagctccaatgccaagcgcccggtctcaagaggtctaaagctcaagtgctcaaccaccagct	2289
Db	2436	CACACCACAGTACCGATGATCCAGTACCAAGCAAGCAAAAGCCAAAGTCCCAACCCACGACA	2495
QY	2290	ccgtctgaataacaagaccgcgaagtgtcgcgaacaactggaactctggaagctctatgag	2349
Db	2496	CCAGTAAATATATAAACTGAAATGTGTTCCAAATTATGATTAATCTTGAAAAATTTATATCAA	2555
QY	2350	ttcctgaataatacttaacatctgtgccacaatatatctcgtctctcagcaactatgac	2409
Db	2556	TTTTTAATATCTCATATATATATGTCCAAAATATATTTTGTTGCTCACTCAACTACATATGAC	2615
QY	2410	gagagaatctctaaacagatacaagaataaaccaagaagaagagagaaactgtcctctgt	2469
Db	2616	GAAGAATATTAATAACATATTAATAATTTACAAAAGAGAAAGAAAGCAATTTAAGTTCATGT	2675
QY	2470	gattccactggaacgctgtctcaatctacatccagaacaacatctccggtatgtatctatgtc	2529
Db	2676	GATCCATTAACATTTATGTTTATATATACAAAATTAACATCTCGTAAATGTATTTATGTTT	2735
QY	2530	gattgcctcaagc-----attct	2547
Db	2736	GATAGCTTTAAACATAGTTTATTCACAACTATTATAGGGTTTATGAAAAAGAAATTTGGTTAT	2795
QY	2548	ctctctcaactgtctcatgagataatactgaagaagagatgctgcgaacctgtataaactc	2607
Db	2796	GTATTTATTTATTAAGGAAATTTATGAAAAAGAAATGTTTGTAAATTTATATAAATT	2855
QY	2608	aaagaacaagacagaatlaaagaactctctcgagagaagcttaagaaggtctccactcgtct	2667
Db	2856	AAGATATATGACAAATATTAATAATTTATATGAGGAAGCAAAAAAGTATTCACATCTGTATA	2915

QY	2668	aaactctcttccagctcccaigcaacacgtctctctcaaccttaagaacacccgaa	272	22
Dd	2916	AAAACTCTTTCAAGTTCATCATGTCACCACTTTATTCATTAAACCTTCAGATTAACCCGAA	297	55
QY	2728	gtgagcgttaacagcagcaactctcaactgcaaccaacttaactcaactgaactgttt	278	55
Dd	2976	GTAACTGCAAAATGATGATACATCACTTACAAATTTGAATTAATGTTTAAATTAATTT	303	55
QY	2788	gagaaacccctgtctctccgcaagaataaagaacatctaccagaacttactgacagaa	284	77
Dd	3036	GAAAACATATTGACTCTTTGGAAAAACAAAATAATATACCAGAAATTAATAGTCAAAA	309	99
QY	2848	tcgtccgcaactctcaagaaagatactgaaagacagcgacacattcatacagagac	290	77
Dd	3096	AGTAGTGAAACTTTATGAAAAGATATTTAAAGATAGGATACATTTATTAATGAATCT	315	55
QY	2908	ttcctactctcgfgaaatctaaagccgatataactctcttcaagatgaacttaa	296	77
Dd	3156	TTTACAAATTTTGTAAAAATCTAAAGCTGATGATTAATTAATTCATGAAATGATCAATCAAA	321	55
QY	2968	cgtaagaagctggaagagagacataaagctgaagaaagacactggaactgagctcgac	302	77
Dd	3216	ACGAAGAATATTGAGAGAGATATTAAATTAATTAATAAAACCTTTACAGTTATCATTTGAT	327	55
QY	3028	ctgtacaacaagtacaactgaaacttgagagactcttcgacaagagaagagacgtccgc	308	77
Dd	3276	TTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	333	55
QY	3088	aagataagatgcagatccaagaagttgacctctgtctcaagagagagcttgaagcaactc	314	77
Dd	3336	AAATTAATAAATGCAAAATTAATAAACTTACTTTTATAAAGACAAATTAATAACAAATTTG	339	55
QY	3148	aactcagyaacaaaccgaacacgtatctgcaagactcttcagttgtcttccacaagag	320	77
Dd	3396	AATTCACTTAATTAACCAAGACATGATTTCAAAACCTTTCTGTCTTCTTTAAACAAAAA	345	55
QY	3208	aaggaagccgaagatcgcccgagacaagaaacactctggaagacaacaaagattctctcaa	326	77
Dd	3456	AAAGAGCTGAAATAGCGAAGAACTGAAACACATTTAGAAAACAAAAATTAATTAATGAAA	351	55
QY	3268	cactacaagaagccctgcgcgaagtatttaatgagagatcttccctcgtgaagactctccc	332	77
Dd	3516	CATTATAAGGACTTGTTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	357	55
QY	3328	gagagagacatccagaccgagaaataactacgcagagctcgagaactcaagttcctgtct	338	77
Dd	3576	GAAGATATTAATTCAAACAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	363	55
QY	3388	aagctcgaagcgaagctgaaagagacaactgaaacttggaagaagaagagcctcagactctc	344	77
Dd	3636	AAATTAGAGGAAATTAATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	369	55
QY	3448	ttctagcgaactgcgtctcactcgtatccgcgcgactcaagaaagtcatttaagaacaagaactac	350	77
Dd	3696	TCAAAGGTTTAACTTCATTTTAATTTGCTGAATTAAGAAAGATTAATTAATAATTAATTAAT	375	55
QY	3508	accggaactagcccaagccggaataataaagacggttaatacgcgacttgaatctttaag	356	77
Dd	3756	ACAGGTATTTCTCCAAACGCTTAATTAATTAACGATGTTTAACAATATGCTTAAGATCTTACAA	381	55
QY	3568	aagttcctgccttgaaagaaacagatgttcgcgaactgtgtgtcttgaaactctgcctccgacaca	362	77
Dd	3816	AAATTTTCTCCAGAGGAAACAGATGTTTGCACAGTTGTAAAGTGAAGAGAGATGCGACACA	387	55
QY	3628	ctgagcaggtctcaactaagaagacctgtcatctactcatgttcgagccgagttccaataca	368	77
Dd	3876	TTAGACAAATCTCAACCAAGAAACACGATCAATCAATGATGAGGACAGAGTCTTAACACA	393	55
QY	3688	attcaccacatctgaagcctcgacgcttgaggtctgataagcatcatatgttgctactctc	374	77
Dd	3936	ATTAACATCATCAAAATGTCTGATGATGAAGTATGATGACGATATCATATGATCTATTAATTT	399	55
QY	3748	ggcagagccgaagagacatcagatgcacttcgcgcgacgltgtaaccgltgaagcttcaact	380	77

D	3996	GGAGAAATCCGAAGAAGATTTATGATGATTTAGACGACAGTAGTAAACGAGAGACCACTAACT	4055
O	3808	ccctccgtgatgtataacatctctgtccaaaatcgagaaacgaatcacgaagtgctctactcg	3867
D	4056	ACTTCCGTAATTTGATTAACATACCTTTCTAAATTTGAAAAAGAAATACAGGTTTATATTTA	4115
O	3868	aaacctcgtgaaagcgctctctaaagtcctcgaagaacacgtctgagataacgttggctgac	3927
D	4116	AAACCTTTAGCAGGCTGTTTATAGAACTTTAAAAAAACAATTGAAAAATACCTTTATGACA	4175
O	3928	ctcaatgtcaacgctgtaagaagacatctctgaacgcgcctttaaagaagaaaattlccaa	3987
D	4176	TTTTAATGTTAATCTTTAAGATATTTTAAATTTCCAGTTTAAATTAACGTAATAATTTCAA	4235
O	3988	aagctcttgagagcgcgctctgattccctctaaagacctgacctctctcaactgaagtgc	4047
D	4236	AATGTTTGAATCAGATTTTAATTTCCATTAATAAGATTTAAACATCAAGTAATATCTTCTC	4295
O	4048	aaggaccatatacaagctctccataaagaagaaggaataaattctgtcactgaatcac	4107
D	4296	AAAGATCCATTAATAATTTCTTAATTAAGAAAAAAGAGATAATTTCTTAAGCAGTTATAT	4355
O	4108	tatctcaaggaactccalcgaaccgcgatcaatcttcogtaatgatgtgcgtgggaltac	4167
D	4356	TATATTTAAGGATTTCAATATGATGATGATTAATAATTTTGCAAAATGAGATCTTGATATAT	4415
O	4168	aagctccctgagcgaaaataaacaagctctgcactgcactatataaagaatatacaagat	4227
D	4416	AAATATATATCCGAAANATTTAATACATTTAGATTCAATTAATAAATAATATACACAC	4475
O	4228	aagcaagcgagaaatgaanaataatctgcacctctctgaataacatcgaacccgtctaac	4287
D	4476	AAACAGGTGAAATGAGAAATACCTTCCCTTTTAAACAAATTTGAGACCTTATATATA	4535
O	4288	acaatgagacgaaaaatcgacccctcttcgtaattcactgcctgagagccaagtgctcaact	4347
D	4536	ACAGTTATGATTAATAATTTGATTTATTTGATTTCAATTAAGAACAAAGTTCTTAATAT	4595
O	4348	acttaccgagaagacaaatgtgagaagttaaatcaagaagcctgaactaccataaacaac	4407
D	4596	ACATATGGAATCAAAACGTAGAAAGTTAAATAAAGAATTAATTACTTAAAAACAATT	4655
O	4408	caagacaagctgtgcagatttcaaagaaaataacaattctgcgaaattgcgaacctgctc	4467
D	4656	CAAAACAATTTGGCAGATTTTAAAAAAATAACAATTTGTTGGATTTGCTGATTTATCA	4715
O	4468	accgatataacaacaacaaatctccgcgaacgaattctgtcaactcggaaatgtgtcgaa	4527
D	4716	ACAAATTTATAACCATATATACTTATTGACAAAGTTCCTTAGTACAGGATATGTTTTGAA	4775
O	4538	aaectcgcaaaaacagtgctgagcaatctgtctga-cggcaacctgcgaagcatgtgaa	4586
D	4776	AACTTGTCTAAATTCGGTTTATTTATTAATTTACTTGATTTGGAACCTTGCAAGTATGTTAA	4835
O	4587	catctccagacacaatgctgtgaagaacaag--cccacgaatagagtggtttcaag	4644
D	4836	CATTTCACAAACCCATGCGCTTAATAAATAAATGATCCAAACAGATTCTGATGTTTCA	4895
O	4645	catctgacgagcgcgaaagtgtaagtgctctccctgaactaacaacaagaagaagataag	4704
D	4896	CATTTAGATGAAGAGAGAATTAATGTTTTTAATTAATTAACAAACAGAAAGTATGTA	4955
O	4705	tgggtgagaagaccaaaacctactctgcgaatgaagaacaatgctgcggtgtgtgacccgatgtc	4764
D	4956	TGTTTAAAAATTTCAATCTCTACTTTCTTAACGAAATTAATGCGTGCTGATGCGAGATGCC	5015
O	4765	aaatgcaccgaggaagacacgcgtctcaacggaagaagaataacaatgagtgtaactaag	4824
D	5016	AAATGTTACGGAAGAAGATTTCAGGTACCAACGGAAGAAAAATATCATGTTGAATGTACTTAA	5075
O	4825	ccgaactctataccactctcgacgggaatttttgcctcagctcaattcttcctggatc	4884

[illegible]

RESULT 7

Accession	LOCUS	DEFINITION	Accession	LOCUS	DEFINITION
E00656	E00656	5760 bp RNA P195 protein of plasmodium falciparum.	E00656	E00656	5760 bp RNA P195 protein of plasmodium falciparum.

ACCESSION	E00656
VERSION	E00656.1
KEYWORDS	GI:2168935
SOURCE	JP 1986019490-A/1.
	JP 1986019490-A/1.

SOURCE	ORGANISM	REFERENCE
Plasmodium falciparum.	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	1 (bases 1 to 5760)

AUTHORS Ansoniil,A.H., Malkuru,J.R., Jiyasublaa,S.S., Barentelna,R.M. and Kareu,G.O.

TITLE DNA CLONING FOR ANTIGEN OF PROTOZOA

JOURNAL Patent: JP 1986019490-A 1 28-JAN-1986;

WELLCOME FOUND LTD:THE
OS Plasmodium falciparum
PN JP 1986019490-A/1

PD 28-JAN-1986
PE 22-FEB-1985 JP 1985034280
PR 22-FEB-1984 GB 84 8404692, 26-SEP-1984 GB 84 842440 PI
ANSONIT AASNA HORUDGA, MAIKURU JIEMOSU ROTSUKUYAA, PI
JYASUBIDAA SHINGU SANDEVU, BARNTEFINA RIBEROSU MOENENO, PI
GERISUTRO ODEIKU
PC C12N15/00,A61K39/015,C07H21/04,C07K13/00,C12N1/20,C12P21/02,
PC C12N1/20,
PC C12R1/19),(C12P21/02,C12R1:19);

```
CC strandedness: Single;
CC topology: Linear;
CC hypothetical: No;
CC nt-sequence: No;
```

CC	anti-sense: NO;	
FH	Key	Location/Qualifiers
FH		
FT		
CDS		216 5180

	FT	FT	FT	FT	FT
sig_peptide	219	.272			
mat_peptide	273	.5177			
	/product='P_195	protein'			

FEATURES	Location/Qualifiers
source	1. .5760

```

/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
BASE COUNT      2565 a      630 c      725 g      1840 t
ORIGIN

```

Query Match	54.8%	Score 2706.2	DB 10	Length 5760
-------------	-------	--------------	-------	-------------

	Best Local	Similarity	/2.1%;	Pred.	NO. 0;
	Matches	3587;	Conservative	0;	Mismatches 1343;
					Indels 45; Gap

by 10 atgaataacatcttccctcgttcattctcgttttatcaaccaaatcagtcgctg
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 216 ATGAGCATCAATCTTTTATGTTCATTCTTTTTTAAATATAACACAATGTGTA 2

QY 70 acccaagcaatcctatcagagctggttaagaaactygaagctttggaagatgcgctcctt 129

130 acccgatacagcctgttcacagaagagagatgtgctcgtaatgaaggagacgagtgcgcaag 189

Db	336	ACAGGTTAAGTTTATTTCAAAAGGAAAAAATGSTATTTAATGAAGGACACAGTGGACA	395
Qy	190	gccgltacaacacgacacacccqgltcctaagqgctgctgacgaqgctgcgcgqtgqg	245

Db 396 GCCTTACAACTAGTACACCTGGTTCAAAAGGGTTCAGTTGCTTCAAGTGGTTCAGGTGGC 455

Oy	250	tcgtagccctcgggggtctccgcgcctccgcgcggaagctggactaaggctggctcaag	309
Db	456	TCACGTCTCTTCAGGTCGGCTCAGTTGGCTTGAGGTGGCTCCAGTTGGCTCCAGCTT	515
Oy	310	gcaagcgcggttcgccggaacagtcgaagaaccaatccactcgaacaactcagcgatcc	369
Db	516	GCTTCAGGTGGTTAGGTTAATTCACACGTCACAAATCCTTCAGATATATTCAGATGATTC	575
Oy	370	gaagcgaagtcctcgaagcgcgcctcaagcgcgagctgagagaactatcctcactcaag	429
Db	576	GATGCTAAATCTTTACGCTGATTTTAAAAACAGAGTCAGAAATTTACTTGTTAACTATCAA	635
Oy	430	gaagctgaagtaaccacagctgtgtcgcacctcaactatcatatgctgcaactgtgataac	489
Db	636	GAACCTCAAAATCCTCAACCTTTGGATTTAACATAATCAATGTTAACTTGGTGATTAAT	695
Oy	490	atcatgctccaataatcctgacttgcgcgtttcaagagagatcaatgaaactccctgacag	549
Db	696	ATTCAATGGTTTCAAATATTTTAATTGATGCTATGAGAAATTAATGAATTAATATATAA	755
Oy	550	ttgaattctactcgcgcgtctgactgaagggccaaactgaaatgacgtttgcgcactgact	609
Db	756	TTAAACCTTTATTTTGATTATTAACAGCAAAATTAATGATGTATGTCTAATGATTAT	815
Oy	610	tgctaaattccatccaatttgaagatcgaagccacagctgtgacgttatgagaagctg	669
Db	816	TGTCMAATACCTTTCAATCTTAAATTTCCGCAAAATTAAGATTTGAGCTCTTAAAAACCTT	875
Oy	670	gtctcgatctgcgaagcctctcgcacacatcaagaagaatgctggagaagctgaagat	729
Db	876	GTGTTCCGATATACAAAACCATTAAGCAATATTTAAAGATATGTGAGAAAATGGAAT	935
Oy	730	tatatlaaagaagaaagaaacccatcgcgaacatcaagcgtgatcgaagatcccaaa	789
Db	936	TACATTTAAAAAAATAAAAAAACATAGAATAATTAATGAATTAATGAAGAAGTAAAG	995
Oy	790	aagaccatagacaaaataaagaatgcacccaaggaagaaagaagaagatgttaccag	849
Db	996	AAAACAATTGATTAATAAATAAAGAAATGCACTTAAAGAAAGAAAAAAATTAATACCA	1055
Oy	850	gccagtagacccgcgtccatcatlaacaagaagctggaagaagccatgaactcaagc	909
Db	1056	GCTCAATATGATCTTCTATTTTCAATTAANCAATTAAGAAAGCACATTAATTAATTAAGC	1115
Oy	910	gtacttggagaagcgcatalagaccctcgaagaagatgaaatcatcaagaagctgcac	969
Db	1116	GTTTTAGAAAAACGTATTTGACACTTTAAAAAAAATGAAGAACATTAAAGSAATTAAGT	1175
Oy	970	aagatlaatgaaltlaagaatcctccgcgaagccaaactctggagacacccctaaacgctg	1029
Db	1176	AAGATTAATGAAATTAATAAATCCGCCACCGCCAAATTCGAAATTAACCAAAATCTCTC	1235
Oy	1030	ctggaacaagaagaagaatagaagggcgccggaagaagatacaagaagatcgcgaaac	1089
Db	1236	CTTCAATAGAACAAAAAATCGGGAACACGAAAAAGAAATTAAGAAATTTGCCAAACT	1295
Oy	1090	attaagctcaacalagatctcctcttactgactcccttgagctgagctactacttga	1149
Db	1296	ATTAAATTTAATNTGATTAATTTAATTTAGTGATCCCTGATTAAGTAATACATTTAAGA	1355
Oy	1150	gagaagaataagaatagacatcctccgcgaagtcgagacacaagaagatcaacgcgaact	1209
Db	1356	GAAAAAATTAATAATATTGATTAATAGTCGAAGAGGTGAANAACAAAGAAATCAACTGACCC	1415
Oy	1210	aatgaatatccaatgtgtgacgttaccctctgctcttataagaagatatacaaaagctctc	1269
Db	1416	AATGAATATCCAAATGGAGTTACTTATCCCTTGTCTATTAACAGATTAATTAACATGCTTTA	1475
Oy	1270	aagagctccaatagctctcgttgcactgataaacccttgactatcagagaagaaccctt	1329
Db	1476	AATGAATTAATTTCTTTTGGTGATTTAATTAATTAATCCATTGATTAATACAAAGAACCAAGT	1535
Oy	1330	aagaatatctacaagacaaatgagagaagaagttatatacagaatatcaagagaagatc	1389

Db	1356	AAAAACATATTA	CTGATTAATGGAAGAAAAAAATTCATTAATGAAGAAATTAAGAAAAATTT	15050
Oy	1390	aaaaatgagaagaagaanaattgagagtgacaagaanaaagttacgaagaacccgacgaagaat	1449	
Db	1596	AAAAATGAAAAA	AAAAAATTTGGAATCTGATTAATAAAAAATCTTAGCAAGACAGATCTAAGCT	16555
Oy	1450	ctaaacgatcactcaaaagagta	tgaagaagctgctgaacgagatcatalgatccaaatlc	15090
Db	1656	TTAATATGATATA	ACAAAGAAATATGAAAAATTTCTTAATGAAGAAATTTAATGATAGCAAAATTC	17150
Oy	1510	aacataacatcgacc	ccgaccacttcgagaaatgatatgggaaacgylactcttacaa	15069
Db	1716	AATATATATATAG	TTTTAACTAATTTTCGAAAAAATGATGGTAAGAAAGATATTCATATATAA	17750
Oy	1570	gtggagaacactgaacacaa	taataactctgtgacaaatcttcaagaatcttcaagaataactt	16290
Db	1776	GTGGAAGAACTTAC	ACACCTTAATTAATTTGCTATCTGCTATGAATAATTTCTTAACATATCTT	18355
Oy	1630	gagaagctcacccaaagct	ctcttaagatalgagagacatcttcctcggaacatigtgtg	16690
Db	1836	GAAAGATTAA	CAAAAGCTCTTAATTAATATGGAAGATTAATTTCTTAAGGAATTAAGTACTGTT	18950
Oy	1690	gagagaagaactcaag	tattcttacaagaatctcataagtaagaatcgaaacgagatcgagacg	17490
Db	1896	GAAAAAGATTTAAAT	TTTTTAATAATTTTAATTAAGCAAAATGAAAAATGAGATTTGAAAAA	19550
Oy	1750	ctctgtgagaacat	taagaagagtgagaacagctgttggagaagaagatcacaaagac	18090
Db	1956	TTAAGTTAAAAAT	TTTAAAAAAGATGAGAGAACGCTTTTGAAAAAAAATTTCTTAAGAC	20150
Oy	1810	gaaataaaccaga	tgaagaagctccgtaggctcccgatattgttaagttccaaagtcag	18690
Db	2016	GAAATATAAC	CACAGATGAAAAAATTTTGAAGATGATCTGACATTTGTAAGATACAACTTCAA	20750
Oy	1870	aaggtgctctcat	gaaacaaagattgatgaactcaagaagactccaactcatcttgaagaac	19290
Db	2076	AAAGTTTATTAAT	TAATGAACAAATTTGACGAATTTAAAAAACATCAATTTGATTTTAAAAAAT	21350
Oy	1930	gtggagttttaa	aaactataatactatctgtgcccgaatagttaagccagagataaagcagaa	19890
Db	2136	GTAAGATTTAAAC	TATTAATATATACATGTTCCCAATTTCTTAACAAACAGAAATTAAGCAAGAA	21950
Oy	1990	ccatactacatc	atcgtactcaagaagaagatagacaaacttgaagttlctatgccaa	20490
Db	2196	CCTTATTAATTA	TAATTTGCTTGAAAAAAGAAATTGATTAATTTAAAGTGTTCATGCTCAAG	22550
Oy	2050	gtcgaagacctga	tcaacgagaagaagaacatlaaacttgaagagacgtcagataac	21090
Db	2256	GTAACATCAT	TGATTAATTAATGAGAAAAAATAATTAACCAAGCAAGTCATCGCATTAAT	23150
Oy	2110	tcgagaaccttca	agaagaagagataaaccgagcaggtctaccacaaagcccggaagaacag	21690
Db	2316	TGCGAACCAT	ATACCCGAGAGAAATTAACAGCAACATCTACAAACCTGGAACAAACA	23750
Oy	2170	gcgcgttcaagct	ctcgcgaagcgatagcgtgtgcaagctccaagcacaagacagagcagaagcga	22290
Db	2376	GCAGAGTCTG	TTTAGAAGAGATTTCTGATACAAAGCAACAGCAACGAACAAACAAACACA	24350
Oy	2230	cagcctccagtg	ccagtgccgtlctcagaaggtcgaagctcaagtgcttaccaccacagct	22890
Db	2436	CAACACACAG	TACACAGTACACAGTACCAAGCAAGCAAGCAAGCAAGTCCCAACACACACACA	24950
Oy	2290	ccgttgaataaca	agaccgggaatgtcagcgaactgacttactcttgagaagctcatalgag	23490
Db	2496	CCAGTAATAT	ATAAATCTGAAAAATGTTTCCAAATTAATGATTAATCTGAAAAATTAATATCA	25550
Oy	2350	tctcgtgaataca	tccatactctgcacaaatataatccgtctcaccagacatatgac	24090
Db	2556	TTTTTAATTA	CTTCAATATATATGTCACAAATATATTTTGCTTCACTCAACATATAGAC	26150
Oy	2410	gagaagattc	ttaacagtaacaaglaaaccaagaagagagtaaacgtgcbctctgt	24690

Db 2616 GAAAGATATTTAAACATATATAATTTACAAAGGAGGAAGAAACAAATTAAGTTCATGT 2675
QY 2470 gttccactgacctgctgttcaatcaccagaacacattccgttatgtatctatgctc 2529
Db 2676 GATCCATTGACTTATGTATTAATATACAAATAATACATACCTGTAATGTATGTAGTGT 2735
QY 2530 gatagcctcaaca-----attct 2547
Db 2736 GATAGCTTAAACATAGTATTATATCACAACATATTTATGGTTTATGAAAAGAAATGGTTAT 2795
QY 2548 cctctccactgtctcctgtagagataatgagaagagatgctgcgaacctgataaactc 2607
Db 2796 GATATTAATATTAATCTTATGAAAAATTTGAAAAAAGAAATGGTTTATTAATTAATTAACCT 2855
QY 2608 aagacaagaagaataatgaacacctcgtggaagaagctgaagaagctccacctctgtt 2667
Db 2856 AAGGATTAATGACAAATTTAAATTTATTAAGAGGACGAAAAAAGATATCCATCTGT 2915
QY 2668 aaaactctctctccagctccatcgaacacactgctctcaccacctcaagaacacccga 2727
Db 2916 AAAACTCTTCAAGTTTCATCAATGCAACCATTTATCACTTAACACCTCAGATTAACCCGAA 2975
QY 2728 gtgagcgctcaagaagacacctctcaactcgaccaaccttaatactcaactgaactgtt 2787
Db 2976 GTAAGTGCAAAATGATATCATCATCATCTACAAATTTGAAATTAATGTTTAAATTAATTT 3035
QY 2788 gagaacatctctctcgtcgcaagaataaagaacatctaccagaactcttctgcaaga 2847
Db 3036 GAAACATATTTGAGCTTGTGAAAAAAACAAATATATACCAACAATTAATAGTCAAAA 3095
QY 2848 tgcgcgagaactctcagagaagatactgaagaagacagacacatctcctaagaagag 2907
Db 3096 AGTAGTGAACCTTTATGAAAAAGATTTAAAGATGATGATCATTTTATTAATGATCT 3155
QY 2908 ttccactaactcgtgtaaatctaaagccgatgtatcaactctctcttaagatgaactaa 2967
Db 3156 TTTACAAATTTGTAAATCTAAAGCTGATGATTAATTAATTCATGTAATGATGATGATG 3215
QY 2968 cgtlaagaagctggaagagacatcaatgaagcgaagacactctgaactctgaactctgac 3027
Db 3216 AGGAACAATTTGAAGACATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3275
QY 3028 ctgtcaacaaglacaaactctgaagactctggaagactctcgaacaagaagaagacgtcg 3087
Db 3276 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3335
QY 3088 aagtataagaatcagaatcagaagtgactctgctcaagaagagactctgaagaacac 3147
Db 3336 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3395
QY 3148 aactcactgaacatcgaacacagctcgaactctcgaactctcgaactctcgaactctcga 3207
Db 3396 AATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3455
QY 3208 aaggaagccgagatcgcgagagagagaacactctggaagaacacgaactctctca 3267
Db 3456 AAGGAAGCTGAATGCAAACTGAAACACATTTGAAACACAAATTAATTAATTAATTAAT 3515
QY 3268 caactcaagaagctcgtcgaatataatgaatgaatcgtcctccctctgaagctcctcc 3327
Db 3516 CATTTAAAGGACTTGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3575
QY 3328 gaggagagacatcagacagagataactcagcagcctcgaactctcgaactctcgaactct 3387
Db 3576 GAGGATTCATTCACAAACAGATATTAATGTCAGTTTGAAGAACTTTAAAGATTAATTAAT 3635
QY 3388 aagctcgaagcagactcgaagacacactggaactctggaagaagaagaagctcgaactc 3447
Db 3636 AATTAATGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3695
QY 3448 tctagcgagactcactcactgactcgcgagctcgaagaagatcatttaagaacaagaactc 3507
Db 3696 TCAAGGCTTTACATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3755

QY 3508 accgcaatagcccaagcgagaaataacagacgtgaataaagcactcgaactcttaag 3567
Db 3756 ACAGTAATTTCTCCAGGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3815
QY 3568 aagtlcctcgtcgaagaagacagatcgcgcactggtgtctcgaactcgtcgcgcgcgcgc 3627
Db 3816 AATTTCTCCAGGAGCAAGATGTTGCAACAGTTGTAAGGAAAGTGATGATGATGATGAT 3875
QY 3628 ctggagcagtlcctaactaagaacctcgaactcactcactcactcactcactcactcact 3687
Db 3876 TTAGAACAAGGTCACCAAGAAACAGATCAACATCAATGATGAGACAGATGATGATGAT 3935
QY 3688 attacacatcctcaagaagctgacagatgaggtcgaatgaactcactcactcactcact 3747
Db 3936 ATTAACAATTCACCAAAATTTCCATGATGATGATGATGATGATGATGATGATGATGAT 3995
QY 3748 ggcgagagcgaagagactcagatgacacctcgccaggtgtcaccggtgagcgtcgaact 3807
Db 3996 GGAGATCCGAGAGAGATTAATGATGATTAAGCAAGTATGTAACAGGACAGCAGTAAT 4055
QY 3808 ccttcgtgattgataactctcgtccaaaatcgagaacgaataagagtgctcactcgt 3867
Db 4056 ACTTCGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4115
QY 3868 aaacctcgtcgcagcgtcctatagctcctcaagaacacactcgaagaatagctgagc 3927
Db 4116 AATCCTTTGCGAGGTGTTTATGAACTTTAAACAAATTTGAAATTAATTAATTAATTAAT 4175
QY 3928 ttcactcgaagcgtgagaagacatctgaacacgcctcttaataagaagaagaactcga 3987
Db 4176 TTTAATGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4235
QY 3988 aagctcttggaagagactctcctccttaagaagcctgaactcctcactcactcactc 4047
Db 4236 AATGTTTGAATTCAGATTTAATTCATTAATTAATTAATTAATTAATTAATTAATTAAT 4295
QY 4048 aaggaaccatacaagtlcctcaataagaagaagaggaataaactcgtcgtactcaaac 4107
Db 4296 AAGATCCATTAATTAATTTCTTAATTAAGAAAAAGAGATTAATTTCTTAAGCAGTTATTAAT 4355
QY 4108 tatatcaagactcactcagacacagatalcaactctcgtatgaatgagtgctcgtggaat 4167
Db 4356 TATATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4415
QY 4168 aagatcctcgaagcaaaatacaagctcgtcgtcgtcctcacttaaaagaatcgaact 4227
Db 4416 AATATTAATTCGGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4475
QY 4228 aagcaagcgaagaatgaaataatctgcctcctcgtgaataacatcgaacccctgtaaac 4287
Db 4476 AAGCAAGGTAATGAGAAATACCTCCCTTTTAATTAATTAATTAATTAATTAATTAATTAAT 4335
QY 4288 acagtgacgacaataatcgaactcctcgtatcaactcgtggaagcgaagctcactcaat 4347
Db 4536 ACAGTTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4395
QY 4348 acttcgagaagaagcaatgaggtgaatgaatgaaggtcgaactcactcactcaaacatc 4407
Db 4596 ACATTAATGAGAAATCAACCTAGAAAGTTAAATTAATTAATTAATTAATTAATTAATTAAT 4655
QY 4408 caagaacagctcgaactctcaagaanaaatacaactcctcgtcgaatcgaactcgtcgt 4467
Db 4656 CAAGCAATTTGCGAGATTTTAAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAAT 4715
QY 4468 accgatltacacacaacaactcctcgtacaaagtlcgtcactcgtcgaatgagtggtcga 4527
Db 4716 ACAGATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4775
QY 4528 aacctcgcgaacagctcgtcgaactcgtcga-ccgcaacccgtcgcgagagagctcga 4586
Db 4776 AACTTGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4835

Dh 2685 AAGATATCCACATCTGTAAAAAAGCTTTTCAGTTCATCAATGCAACCATTAATATTAACA 2744
Qy 2710 cctcaagacaagccggaatgagcgctaaacgacacgtctcaactcgacacccctaa 2769
Dh 2745 CCTCAGGATTAACCCGAGTAAGTGCATAATGATACATTCATTTCAAAATTTGAAT 2804
Qy 2770 aactcactgaaactgtltgagaacatccgtctcgtcgcgaagaataaagacatctacaa 2829
Dh 2805 AATAGTTTAAATTTATTTGAAAAACATATGAGCTTTGGAAAAACAAATATATATCCAA 2864
Qy 2830 gaactatttgacagaagaatcgctccgagaacgtctcagagaagatctagaagaacagcgac 2889
Dh 2865 GAATTAATAGTCAAAAAGTAGTAAAACTTTTATGAAAAGATATTTAAAGATAGTGAT 2924
Qy 2890 acattctaaagagagagcgctcctcaactcgtgaaatctaaagccgagtatactcaact 2949
Dh 2925 ACATTTTAAATGAATCTTTTACAAATTTTGTAAATTCAAAGCTGATATATTAATTCGA 2984
Qy 2950 cttaacgatlgaatctaaacgtlaagaagctggaagagacatcaataagctgaagaagaca 3009
Dh 2985 TTGATGATGATCAAAAAGAGAAATTAGAAGAAAGATTAATTAATTAATAAAAAACT 3044
Qy 3010 ctgcaactgagctcgacgtcgtacaaagtlacaacgtlaaactggaagagactctcgac 3069
Dh 3045 TTACAGTTTCAATTTGATTTATATTAATTAATTAATTAATTAAGAACTTAATTTGAT 3104
Qy 3070 aagaagaagacgctcgagcaagatagaatgacagatcaagaatltgactctgctcaagag 3129
Dh 3105 AAAAAAGAAACGTTGGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3164
Qy 3130 cagctltgaagaacaaactcaactcaactgaacaaatccgaacacgtctcgagaactctca 3189
Dh 3165 CATTAATGATCAAAATTTGATTTCACTTAATTAATCCAAACATGATTAATCAAAAAGTTTCT 3224
Qy 3190 gtgtcctcaacaagaagaagacccgagatcgccgagacagagaacacgtctgagaac 3249
Dh 3225 GTTTCTTTTAAACAAAAAAGAAAGCTGAAATAGCAAAAGCTGAAAGCTGAAAGCTGAAAG 3284
Qy 3250 accaagatctctctcaacaacacacaaagccgtcgaaglatataatgtgagagctctc 3309
Dh 3285 AAAAAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3344
Qy 3310 cctctgaagacgtctcgcgagagagacatccagacgagatcaactcgcaccccgag 3369
Dh 3345 CCATTAATAAACTTTAAGTGAAGATCAATTCAAACAGAGATTAATTAATCCAGTTTAGAA 3404
Qy 3370 aactcaaggtcgtctgaactcgaaggaagctggaaggaacacgtctgagaag 3429
Dh 3405 AACTTTAAAGTTTAAGTAAATTAAGAGAAATTAAGATTAATTAATTAATTAAGAAAG 3464
Qy 3430 aagaagctcagctacgtctcgtcagcgactgacacgtcgtcgcgaggtcagaagac 3489
Dh 3465 AAAAAATTAATCACTTATCAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAAT 3524
Qy 3490 attaagaacaagaactacacggaacatagcccaagcgagaatcaataacagaagtlgaatc 3549
Dh 3525 ATAAAAAATTAATAATTAACAGTAATTCCTCAAGTGAAGATTAATTAATTAATTAATTA 3584
Qy 3550 gcaactggaactctacaagaagtlctcgtcgtcgaagaacagatgctgcgaactgtgtgtc 3609
Dh 3585 GCATTAATGATCTTAACAAAAATTTCTCCGAGAGAACGATGTTCACACAGTTTGAAT 3644
Qy 3610 gaattgtcgtcagacacgtgagagcttcaacctaaagaacgtcgtcgtcgtcgtcgtcgtc 3669
Dh 3645 GAAATGGATCCGACACATTAACAAAGTCAACCAAAAGAACCAACGATCACTCACTGTA 3704
Qy 3670 ggaagcgagatccaataacatccacatctcgaagacgtcgaagatgaggtlctgacgtc 3729
Dh 3705 GGAGCAGAGTCAACACATTAACACATCAACAAATGTGATGATGAATTAATTAATTAATTA 3764
Qy 3730 atcatgtgctatactctcgagagagagagagagacatgacgtccgagaggtgtc 3789
Dh 3765 ATCATAGTACCTATTTTGGAGAAATCCGAAGAAAGATTAATGATGATTTAGACAAAGTGA 3824

Qy 3790 accggtgagcgtcgaactcctccggtatgataacatctgtccaaatcgagaacga 3849
Dh 3825 ACAGGAGAGCGATACCTCTCCGTAATTTGATACATCTTTCAAAATTTGAAAAATGAA 3884
Qy 3850 tacgaagtgctatctgaacacctcgtcagagcgctcatalaggtctctcaagaacagctg 3909
Dh 3885 TATGAGCTTTTATTTAAACCTTTAGCAGGTTTATAGAAGTTTAAAAAACATTA 3944
Qy 3910 gagaataacgtgagactcctcaatgcaagtgaaagagacatctgaaacgacctaa 3969
Dh 3945 GAAATTAACGTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4004
Qy 3970 aagaagaagaatltcaagaagctgtgagagcgactgtatcccaataaagaacgtcacc 4029
Dh 4005 AAGCGTAAATTTCAAAATGTTTGAATGATGATTAATTAATTAATTAATTAATTAATTA 4064
Qy 4030 tccctcaactacgtltgcaagaagccatatacaagtlcctcaataaagaagaagagataa 4089
Dh 4065 TCAAGTAATTAATTTGTTCAAGATCCATTAATTAATTTCTTAATTAAGAAAAAGAGTAA 4124
Qy 4090 ttctgtctagtltaacatatacaagaactcctacgcacacgataatcaatltcgctaat 4149
Dh 4125 TTCTTAAGCAGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4184
Qy 4150 gatgtgtggtlatlacagaagctccgagcgaaataaagaatcgtcgtcgtcgtcgtcgtc 4209
Dh 4185 GATGTTCTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4244
Qy 4210 aaaaaglatatacaagataagaagcgagaatgaaataatctgcctcctcgtataac 4269
Dh 4245 AAAAAATATATACAGCAGCAACAGAGTGAATTAATTAATTAATTAATTAATTAATTAAT 4304
Qy 4270 atcgaaacctgttacaagaagtlgaacgaacaaatcgacacctcgtglaattcaactgag 4329
Dh 4305 ATTGAGACCTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4364
Qy 4330 gccaaagctcctaactatacttaagaagaagagcaatltggaagtttaaatcaagaagcgt 4389
Dh 4365 GCAAAAAGTTCTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4424
Qy 4390 aactacctaacaacatccaaagaagcgtgcagatltcaagaagaataacatltcgtc 4449
Dh 4425 AATTAATTAATAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4484
Qy 4450 ggaattgcagacctgtctacgattataacgaacaaatcctcgtacaaagtlctcgtc 4509
Dh 4485 GGAATTTGCTGATTTATCAACAGATTAATTAATTAATTAATTAATTAATTAATTAAT 4544
Qy 4510 acttgcatgtgttgcgaagaacctgcgaagaacagtgagagacatcgtctgcagagac 4569
Dh 4545 ACAGGATGTTTGAATTAATCTGTAACCGTTTATTAATTAATTAATTAATTAATTAATTA 4604
Qy 4570 ctgcagggatcgtgaacatcccaagcacaatgcgtlgaagaagaagtgcccccgaat 4629
Dh 4605 TTGCAAGGATGTTTAAACCTTTCAACACACCAATGCGTAAAAAATTAATTAATTAAT 4664
Qy 4630 agcagctgttcaagacatcgtgcgaagcgagaagtgcaagtlctcgtacatacaa 4689
Dh 4665 TCTGATGTTTCAAGCATTTAGATGAAGAGAAATGTAATGTTTATTAATTAATTAATTA 4724
Qy 4690 caagaagaagataagtgcgaggaacccaacacctacgtcaatgtgaagaagaacggtg 4749
Dh 4725 CAAGAGGATGAATTAATGTTTGAATTAATTAATTAATTAATTAATTAATTAATTAAT 4784
Qy 4750 tgtgacgccaatgctlaaatgcaacgagaagaacgagcgtcctcaacgagaagaataacaa 4809
Dh 4785 TGTGATGCGATGCAAAATGTAACGAAGAGATTAATTAATTAATTAATTAATTAATTAAT 4844
Qy 4810 tgcagtgtaacgaagcctcctatacactcctcgaaggaatlttgtctcgaactc 4869
Dh 4845 TGTGATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4904

Dh 906 GCTCAATATATCTTTTATTATTAACAATAAACATTTACAGAAGACACATATTTAATAAGC 965
Qy 910 gtaacggagaacgcgatagacaccctcaagaagaatgaanaatatcaaaactgcgcgac 969
Dh 966 GTTTTAGAAAAACGTTGTACACTTTAAAAAAAATGAAGACCTAAAGAAATTTACTGTGAT 1025
Qy 970 aagattatgaatlaagaactcctccgcagcaactctgggaacaccccttaacagcgtg 1029
Dh 1026 AAGATTAATAAGAAATTTAAAAATCCCGCGGCAATTTCTGGAATATCACCAATATCTCTC 1085
Qy 1030 ctggaacaagaacaagaagatagaagagcaggaanaagagatcaaaagatcgcgcaaac 1089
Dh 1086 CTGTGTAAACAAAAAATTCAGAACGAGAAAGAAAAATTAAGAAATTTGCCAAAACT 1145
Qy 1090 attaatgaacatgatctctcttactgatacccttgagctgagatctacttgaga 1149
Dh 1146 ATTAAATTTAACTGATGATGTTATTTACTGATCCACTTGAAATTAAGAAATTTTAAAGA 1205
Qy 1150 gagaagaataagaatataagacatctccgcgaagtcga-----gacaagaagatca 1200
Dh 1206 GAAAAAATAAATAAGTTGATGTAACTAAATCAAGATCTTACGAATCTGTTCAA 1265
Qy 1201 accgaactaatgaatatacccaatggtgtagcgtacccctcgtctataaagatalcaac 1260
Dh 1266 ATACCAAAAGTCTCTTCAATGCTATTTGATATCTTACCACTCATCATTTTCAT 1325
Qy 1261 aa-----cgctcacaagcagctcaatagcttcggtgagcttgattacccctcgat 1311
Dh 1326 AATTCTATTGCTGCAATATATATAAAATTCATATGATGATTTAAATGAAATCTGTATCT 1385
Qy 1312 taacagaagaacccctctaagatatctacagaacatcgagaagaagaattatacaac 1371
Dh 1386 AAGAAAAAATTAATGAATAAATTTATTACAGATTAATAGGAAGAAATTTATTCATTAAT 1445
Qy 1372 gaatcaagagaagatcaaaatcgagaagaanaaattggaagtgaagaagaattac 1431
Dh 1446 AACATTAAAAAACAATGATTAGTAAGAAAAAACAATTAAATCACAAAGAACAAAT 1505
Qy 1432 ggaagccgagaagaagctlaaagcatalactaagaagatgaagaagctgcgcgaacgag 1491
Dh 1506 AAAAAATTAATCTGAAAGATTATGAAAAATCAAAAAAGATTAAGAAATTTACTGAAAAA 1565
Qy 1492 atctatgattccaatctcacacataacatcagcctgacccaactcgcgagaanaattgagga 1551
Dh 1566 TTTTATGAAATGAATTTAATAATATTTTGAAGAAAGATGCTGATGATTAATATTCAGT 1625
Qy 1552 aaagcgtactcttacaagaatgagaagaactgacacacataactccttgcatactgag 1611
Dh 1626 GCAGAGTATATCATATATGTTGAAAAACAAGATTAATTAATTAATTTTCATCTCTAAT 1685
Qy 1612 aattctagaacataactcgagaagctcaccagaagctctaagatatatgagaactcttct 1671
Dh 1686 AATTCTGATATATATGTTCAAAAAATTAAGAGCTCTTTCATATCTTGAAGTTATTTCT 1745
Qy 1672 ctgcggaacactgctgtagagaagaactaaagattatcaagaactcctcaagtaagatc 1731
Dh 1746 TTAAGAAAAAGAAATTTCTGAAAAAGATTTTAATCATTTATTAATTTGAAAACTGGCCTC 1805
Qy 1732 gaanaagagatcgcgagcgtctgtagaacatlaagaagaagatgaagaacagtgtttgag 1791
Dh 1806 GAAGCTGATATTAATAAATAACACGAAGAAATTAAGAGTAGTGAAGAAACAATAATTTAGAA 1865
Qy 1792 aagaagattacacaagaacgaataataacagatlgagaagaatcctgagagctcccgatatt 1851
Dh 1866 AAAAAATTTTAAAGACTTAACACATTTACGCAAAATG---CTTCTTAAGAAATCTGTGATAT 1922
Qy 1852 gttaagatcccaagtgtagaagctcctccatgaacaaagatgagtaactcgaagaagact 1911
Dh 1923 GTAAATTTCAAGATCAAAAGATTTTATTAATTAATAAATAATAGAGACTTAAGAAAAAGATA 1982
Qy 1912 caactcatctcgaagaacgtgtagtlaaacaataatacatatgctgcgaatgataag 1971
Dh 1983 GAATTAATTTTAAAAAATGCACAACTAAAGAAATAGTATTCATGATCTACCAATATTTATAA 2042

Qy 1972 cagaggaataagaagacacatactactcctatcgtactcaagaagaagatagaacaactg 2031
Dh 2043 CCACAAATAATTAACCAACCATTTATTTAATTTGATTTAAAAAAGATAGTAAATTA 2102
Qy 2032 aaagtgtcatgcccagaatcgagacccctgatalcaagaagaagaagatgaacatlaaact 2091
Dh 2103 AAGAAATTTATACCAAAAGTAAAGACATGTTAAAGAAAGAACCAACTGCTTATCAAGT 2162
Qy 2092 gaaggaagctc-----agataactccgagccttccacagaagaagagatagacggagag 2145
Dh 2163 ATTACCAACACTTTACTTACGACCAACGGAACCAATGAAGATGGGGTCACTCCACACAC 2222
Qy 2146 gctcaccaagcccgagaaacagccggttgagcctctgaagaagatgagctgcgagagct 2205
Dh 2223 ACATTTATCCCAATCAGAGAAACAGAAATGTAACGAAAGAAACAGAAAGACAGAGAAACA 2282
Qy 2206 caagcaagaagcagaagcagc-----acaagctccagtgccagtgccggttcca 2256
Dh 2283 GTAGGACACACAAACAGGTAAACATTAACATTTACCAACCAACCAACCATCCACCAAAA 2342
Qy 2257 gaagctaaagctcgaatgctcctacacacagcctcctgtaataacaagaacggagatc 2316
Dh 2343 GAAGTAAAGTTGTTGAAATTCATTAAGAAATGAAGTAATGACAAATTCACAAAGCCTTG 2402
Qy 2317 agcaaacctgactaccttgagaagcctatgagcttgatcctgaataacatcctacacatgcag 2376
Dh 2403 ACAAAAACAGTTTATCTTAAAGAAATTAAGATGAATTTTAACTTAATCATATATATGTCTAT 2462
Qy 2377 aaatatactcgtcctcctacagcaactatgaagaagaagatcttcaaaagatcaagata 2436
Dh 2463 AATATATATTTTATATCAAACTCTACTATGACCAAAAAATTAATTAAGATATTAATCTT 2522
Qy 2437 accaagaagaagagaglaaactgctccttgatgcacacgtggaacgtgcgttcaatc 2496
Dh 2523 ACTCCAGAAAGAAAAATATGAAATTAATGATGATGATGATGATTAATTAATTAATTAAT 2582
Qy 2497 cagaacaacatccgttatgatatgctgctgagctgaagctcaaacatctcctcctca 2556
Dh 2583 CAAATTAACATCTCTCTATGATTTATGATTAATGATTAATGAAATGATTTTACAACT 2642
Qy 2557 ctgtcatgagagatalatgaagaagatgctgcaacactgatalaactcaagaacaac 2616
Dh 2643 CTCTTTTGAATTAATATCAAAAGGAAATGATTTATTAATTAATTAATTAATTAATTAAT 2702
Qy 2617 gacaagattaagaacctctgaggaagctgaagaagctc----- 2657
Dh 2703 AATCACATCAAAAAATTAATTAAGAGCAAAAAACAATTAACCTGGAACATCTACATGCC 2762
Qy 2658 -----cacctcgttaaacctcctctcagcctc 2687
Dh 2763 AGTCTGGAATAATCAACCGTAATTAATCTGCTCAATTCGCAACTCACAGTAATTCCAAAAC 2822
Qy 2688 catgcaacacactgctcctacacactcaagacaa-----g 2721
Dh 2823 CAACATTAATTAATCACTCTCTACCATTAACCAAAATGGTGTACTATCATCTGTGCTCT 2882
Qy 2722 cccgaagtgaagcctgaagaagcagcctcctacatgacccaacttaataactgaatga 2781
Dh 2883 GCTGTAGTTGAAGAAAGATGATGATCCCTTAACAGTATTTCTATTTGTAACGATTTGAAA 2942
Qy 2782 ctgttgagagaacactcgtcctcctcgcgcaagaataaagaact---ctacaagaactatt 2838
Dh 2943 GGATATTGTATGCTCTTAATCTTGAATAATTAACATTAAGTACTTAATCAATTAACCAAT 3002
Qy 2839 ggaagaanaactgctcgcgagaactctcagaagaatctgaagaagcgcagacatctat 2898
Dh 3003 TCTACAAACAGAGATGAAAAATTTTATGAGAAATTTTAAAAAATATGATACCTATTTT 3062
Qy 2899 aacgagaactcactaactcgtgaaatctaaagccgatgatatacactccttlaagat 2958
Dh 3063 AATGATGATATCAAAACATTCGTAATAATCTAATTCAAAAGTATTTACAGTTTGACC--- 3119

[illegible][illegible]

[illegible]

1084 aaacattgaatcaacatagattctcttactgactcccttgagctggaactac 1143
1128 AAACTATTAAATTTACCTTTACTGATCCACTTCAATTAATTAATTTT 1187
1144 ttgagagaagaataagatalagacatccgcgaagtcga-----gacaaag 1194
1188 TTAAGAGAAAAAATAAAGTTGATGACCTTAATCAAGATCTCTAGAAATCT 1247
1195 gaataacaggaaccgaagaataatggtgagagctgactcccttgctataagct 1254
1248 GTTCAAAATACCAAAAGTCTCTATCCAAATGATTTGATATCTTTACCTCTGAT 1307
1255 atcaacaa-----cgcctcaacgagcccaatagcttgctgtaactgtaacccc 1305
1308 ATTCAATTAATTCATTTAGTCGACATATGATTAATAATTCATTTGATTAATCT 1367
1306 ttcgattacgaagaaccccttaagaatctcaacagaaatgagagaagaagttt 1365
1368 GATACTAAGAAAAAATTAATGAAAAAATTTACAGATTAATAGGAAAAAATATTC 1427
1366 atcaacgaatcaagaagagatcaaaatltgagaagaagaatltgagatgacaagaa 1425
1428 ATTAAATTAACATTAATAAACAATGCTTTAGAGAAAAAACAATTAATCACACAAAAGA 1487
1426 agtaccgaagaccgagcaaaagcttaaacgatatcaactaaagatgataaaagctgctg 1485
1488 CAAAATTAATAATTTCTGGAACATTAATGAAGTCAAAAAGATTTATCAACAATTTACTT 1547
1486 aagagatcatgattcacaatcaacaaatcagaccctgacaactcgagaagatg 1545
1548 GAAAAATTTTATGAATATGAATTTATTAATTTTATGACAAAGATGCTGATATAAATA 1607
1546 atgggaaacgctactcttacaagaatltgagaagaacgagacacacataactccttcac 1605
1608 TTCACTGCAAGATATACATATATATGTTGAAAAACAAGATTAATTAATTAATTTTCATCC 1667
1606 tatgaacttctaagcataactcttgagaagctcaccaaaagctcttaagatata tggagac 1665
1668 TCTAATTAATTTCTGTATTAATTTCTCAAAAATTAATAAGGCTTTCTTCAATCTTGAAGAT 1727
1666 tattcttcgcggaacatctgtgtgtgagaagaactaaagatctcaagaatctcataagt 1725
1728 TATTTCTTAAGAAAAAGAAATTTCTGAAAAAGATTTTAATCATTAATTAATCTTTGAAACT 1787
1726 aagatcgaagaacgagatcgagacgctctgtgtgagaacatcaagaagaatgagaagaagctgtg 1785
1788 GGCCTCGAAGCTGATATTAATAAATTTAACAGAAATTAATAAGATGATGAAAAACAATAAT 1847
1786 ttgagaagaagatctacaagaagacgaataaaccagatgagaagaatccctgagagctcc 1845
1848 CTAGAAAAAATTTTAAAGGACTTACACATCTCAGCAATG---CTTCTTAAAGATCT 1904
1846 gatattgttaagttcaagttgagagagtgctccctcattgaaacaagaattgtaactcaag 1905
1905 GATATTGTAATTAATTAACAAGTACAAAAGTTTATTAATTAATAAATAATACAACTTAATA 1964
1906 aagatcgaacatctgaagaagctgtgagttaaacataatatacatctgtccgagatgt 1965
1965 AAGATAGAAATTTTAAATAAATGCACACATTAATAAAGATAGATTCATGTACCAATATAT 2024
1966 tataagcagaagaataaagcaagaacatactactcctcatctgtaactcaagaagaatagac 2025
2025 TATAAACCAAAAATTAACACAGCATATTTAATTTATGATTAATAAAGAAGATGAT 2084
2026 aaactcgaagaagttcattcagcccaagctgagagcctgatacgaagaagaagaacatt 2085
2085 AAATTAAGAAATTTATACCAAAAGTAAAGACATGTTAAA---GAAAGAACAAAGCTGTCT 2142
2086 aaactcgaagaagctgataactcgaagccttcacagagagagagataaacgagacag 2145
2143 TATCAAGATTAACACACCTTTAGTTGACAGCAAGGAAACACATGAAAGAT-GGGGCTCAC 2201
2146 gctaccacaaagcccgagacaaagcgcgttcagcttcgaagcgatagcgltgaagct 2205

2202 TCCACACACACTTATCCCAATCAGAGAAACAGAAAGAAACAGAAACAGTACGACAC 2261
2206 caagcagaagcagaagcagagcaagcctccagtgccagtgccgttccagagctaa 2265
2262 ACACACACGCTTAACATTAATCATTTACCCACACACACACCTCACCAAAAGATGAAA 2321
2266 gctcaagtgctcaccaccagcctcctgtgaaataaagaacgagaatgttcgaagacg 2325
2322 GTTGTGAAAAATTCATTAACACATTAAGACTTAATGACAAATTCACAAAGCTTGACAAAACA 2381
2326 gaactccttgagaagctctcattagttcctcgtgaatacactcaatctcgcacaaatatactc 2385
2382 GTTTATCTAAGAAATTTAGATGATGATTTTAACATAATCATTAATTAATGATTAATTAAT 2441
2386 ctcgtctcacaagcactatgagaagaagatcttaacagatcaagaataacaaagaa 2445
2442 TTAGATCAAACTCTAGATGAGACCAAAAATTTATGAGATTAATTAATCTTCTCCAGAA 2501
2446 gagagagaaactcgtccctcctgtgatacagcagcgtctgtcaatctcagaacaac 2505
2502 GAAAAAATGAATTAATTAATCATATGATGATCCATTAATTTATTAATTAATTAATTAAT 2561
2506 attccgttatgtatctatgtctcgtatagcctcaacaactcctcctcacaactgtcattg 2565
2562 ATACCTGCTATGATTAATTAATTAATGATTAATGACAAATGATTAATTAATTAATTAAT 2621
2566 gagatataagagaagagatgtctgcgaactgataaactcaagaagaagaagaat 2625
2622 GAATTAATCAAAAGAAATGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2681
2626 aagaacctctcgagag----- 2641
2682 AAAAAATTTTATGAGAGAGCAAAAACAATAACTGCAACATCTATCAATCCAGTCTGGA 2741
2642 -----aagctaaagaagctccactcgtttaaactcctcctccagctccaatgcaac 2642
2742 AATACACACGTAATTAATCTGCTCAATCCGCACTCACAGTAATTTCCAAACCAACATCA 2801
2697 actgctctcacaactcaagaaca-----gcccgagtg 2730
2802 AATGCAATCTCTTACCAATTAACCAAAATGCTAGCTGATCAATCTGCTCTCTGAT 2861
2731 agcgtcgaagcagacacacccctcagcagcaactcaataactcaactgaaactgtgtgag 2790
2862 GAAAGAAAGTCAATGATCCCTTAACAGATTTGCTATTAGTAAGATTTGAAAGTATGTT 2921
2791 aacatcctgtctcgcgaagaataagaacat---ctacaaagaacttatgagacaga 2847
2922 ACTCTCTTAATCTTGAATTAATAAATACTTAAGTCAATCCATTAACATTTCTACACACA 2981
2848 tctgcgagaactctcgaagaagatcaggaagaagcagagacatctcataagagagc 2907
2982 GAGATGAAAAAATTTTATGAGATTAATTAATAAATAATGATTAATTAATTAATGAT 3041
2908 ttcactaaactcgtgaaatctaaagccgagatgatacactccttaacgataactcaaa 2967
3042 ATCAAAACATTTCTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3098
2968 cgtlaagaagctlgagaagagacataaataagctgaagaagacatgcaactgagctcgac 3027
3099 AAAAAATGCAATTAATGAGTAATTAATAAATAATTAATAAGATTTTACAGTTATCAATTTGAT 3158
3028 cgttacaagaatcaaaactgagagagccttcogaagaagaagaagcagctggc 3087
3159 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3218
3088 aagtaagaatcagataagaagaagtgactcgtccaaagcagcgttgaagaagaactc 3147
3219 CAAGCAAAATGCAATTAATTAATAAATACTTATTATTAATAAGACAAATTTGAAATCAAAATTTG 3278
3148 aactcattgaagaactcgaagaacgctactgagaagaactcagtgltctcacaagaag 3207

Db 3279 AATTCATTAATACCCACATTAATGATTAACAAAACCTTTCTGTTTCTTTAAACAAAA 3338
 Qy 3208 aagaagccgagatcgccagagacagagacactctgagaaacaccagaatctctcaaa 3267
 Db 3339 AAAGAAAGCTGAAATAGCGAAACCTGAAACACATTAGAAACACAAAAATATTATTGAAA 3398
 Qy 3268 caclacaagacccctcgtaagatataatggcgagctctctcctcgagaactctcc 3327
 Db 3399 CATTTAAAGGACTGTTAAATATTATATGATGTAATCATCTCCATTAAACCTTTAACT 3458
 Qy 3328 gaggagagacatccagacagagagataactacagccagccctcgagaactcaagtcgtct 3387
 Db 3459 GAAGTATCATTTCAAAAGAAATATATGCAATTTAGAAAATTTAGACTATTAACT 3518
 Qy 3388 aagctcgaaggaagctgaagagacactgaacctgagagaaagaagctgaagctacc 3447
 Db 3519 AAAATAGATGGAAACCTCATGATTAATTTACTTTAGGAAACAAAATTAATCTTTTA 3578
 Qy 3448 tctagcgagctgcatcactgcatcgccgagctcaaggaagctcatttaagacaagaactac 3507
 Db 3579 TCAAGTGGATTACATCATTTATTACTGAAATTAAGAAAGTATATAAATATAATTTAT 3638
 Qy 3508 accggcaatagcccaagcgagagataatcagacgctgaatacgaactggaatcttacaag 3567
 Db 3639 ACAGGTAAATTCACAAAGGAAATATTAAGAAAGTTAACGAAGCTTTAAATCTTACGAA 3698
 Qy 3568 aagtcctcgccgagagagacagatgctgcacactgctggtctc----- 3608
 Db 3699 AATTTTCTCCCAAGAAACAAAGTTACAAAGTTGTAACCTCCACCTCAACAGATGTAAC 3758
 Qy 3609 -----tgatctgctccgacacactgagacgctcgaactcaaga----- 3650
 Db 3759 CCATTCATTAATCTGTAAAGGTAGTGTACTTCAGATGTCACAAAGAAACAAACACAA 3818
 Qy 3651 --gectgcatctactcatgctgagcgagctcgaactcaatacacaactctcaagaagctc 3708
 Db 3819 ATACCAATTCAGCTCTTTATTAACAGATTTACAAAGTGTACTCAATCACAAAAATTAT 3878
 Qy 3709 gacgttgagctgcatgagctcatctgctcctatcttcgagagagcgagagagactac 3768
 Db 3879 GACGAGAGAGATGATTCCTTACTGTTATTAACCATTTTGGGAATCCCAAGATTAATGC 3938
 Qy 3769 gatgaccccgccagctgctacccgctgagctgctcactcctcctgctgctgaactac 3828
 Db 3939 GAATTTTATGATCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3995
 Qy 3829 ctgtcacaatctgagagacagatgctcctcctcctcctcctcctcctcctcctcctc 3888
 Db 3996 CTCTCAGATTTGAAATGAATATGATGATGATGATGATGATGATGATGATGATGAT 4055
 Qy 3889 aggtctctcaagaagacagctgagagaaatgctgagacactcctcctcctcctcctcctc 3948
 Db 4056 AGAAGCTTAAAAAAACAAATGAAATGAAACAAATTAATTAATTAATTAATTAATTAAT 4115
 Qy 3949 attcgaacagcgcttcaataagaagaattcgaagacgctctgagagcgagctg 4008
 Db 4116 ATCTTAATTAATCAGCTTAAAGAAACGAAATATTTCTTAAAGATGATTAATTAATTA 4175
 Qy 4009 attcctcaataagacgctcctcctcctcctcctcctcctcctcctcctcctcctcctc 4068
 Db 4176 ATGCAATTTAAACATATATTCCTCAATGAATGATGATGATGATGATGATGATGATG 4235
 Qy 4069 aataaagaagaagagagaaatctctgctgctgctcctcctcctcctcctcctcctcctc 4128
 Db 4236 AATTCAGAAACAAACAAACATTTTAAAAAGTTACAAATATATTAATAAATACAGTAGAA 4295
 Qy 4129 accgaatcaaatctcgcaatgctgctgctgctgctcctcctcctcctcctcctcctcctc 4188
 Db 4236 AATGATATTAATTTGCAACGAGAAAGTATAGTATATTAAGAAAGTTTATAGCAATAT 4355
 Qy 4189 aagctgacactgactcattataaagatatacaagataag----- 4230
 Db 4356 AAGCATGATTTAGATCAATTAATAAAGTTATCAAAAGAAAGAAAGAGTTCCCATCATCA 4415

Qy 4231 -----caagcgagatgaataat 4251
 Db 4416 CCACCACACACACCTCCGACACGCAAAAACAGACGAAACAAAGAGAAAGTAAGTTC 4475
 Qy 4252 ctgcctctctgaataacatcgaaacctgtacaagacgtaagcgaataatcgacctc 4311
 Db 4476 CTTCATTTTAAACAAACATGAGACCTTATACATTAATTAATTAATTAATTAATTAAT 4535
 Qy 4312 ttcgtaattcactctgagagcgaagtlccctcaactactactaagagaagaagcattgga 4371
 Db 4536 TACTTAATTAATTAATTAAGCAAAAGATTAACGATTTGTAATGTTGAAAACATGACACAT 4595
 Qy 4372 gttaaatcaagagctgtaactcctcaaacatcccaagaagctgagactgcaatgaag 4431
 Db 4596 GTTAAATATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4655
 Qy 4432 aaaaatacaattctgctggaattcgagacctgtctacagatataacacacaactc 4491
 Db 4656 AACACTAACGACTTCGAAACATTAATAAATTTGATTAATGATGATGATGATGATGATG 4715
 Qy 4492 ctgaccaagtlctgtccatgcatgctgctgctgctgctgctgctgctgctgctgctg 4551
 Db 4716 CTGGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4772
 Qy 4552 aatctgctgagagcgaactcgagagcagctgctgctgctgctgctgctgctgctg 4611
 Db 4773 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4832
 Qy 4612 aacagctgccccagagatagcgctgtctcagacatctgagagcgcgcgaagagctgcaag 4671
 Db 4833 AAACATGTCGAAATAATCTGATGTTTTCACACATTGATGATGATGATGATGATGATG 4892
 Qy 4672 tgtctcctgaactcaacaagaagagagataagtgctgagagaaacccaactcctgc 4731
 Db 4893 TCTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4952
 Qy 4732 aatgaagaacatgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 4791
 Db 4953 AACGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5012
 Qy 4792 aacggaagaagaatcacatgctgctgctgctgctgctgctgctgctgctgctgctg 4851
 Db 5013 AACGGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5072
 Qy 4852 atttctgctcagctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 4911
 Db 5073 ATTCTGCGAGTTCCTCTAATCTTAAATATCATTTAATTAATTAATTAATTAAT 5132
 Qy 4912 ctgtacagcttcatctaatagatgctg 4939
 Db 5133 TTATACAGTTTCAATTAATAAATGTAGG 5160

RESULT 11
 AF218248
 LOCUS AF218248 5085 bp DNA INV 24-JAN-2000
 DEFINITION Plasmodium falciparum major merozoite surface antigen (gp195) gene,
 complete cds.
 ACCESSION AF218248
 VERSION AF218248.1 GI:6740099
 KEYWORDS
 SOURCE malarial parasite P. falciparum.
 ORGANISM Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 5085)
 AUTHORS Shan,Z.X., Yu,X.B., Li,X.R., Ma,C.L. and Fang,J.M.
 TITLE Molecular cloning and sequence analysis of major merozoite surface
 antigen(gp195)gene of Plasmodium falciparum isolate FCCL/HN
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 5085)
 AUTHORS Shan,Z.X.
 TITLE Direct Submission

[illegible]

Db	3514	TCAGTGGATTACATCTTTTAATTACTGATTTAAAGAAAGTATATAAAATATAAAATTTAT	3573
Oy	3508	accggcaatagcccaagcsgaaataacagacytaataacgcactygaactctacaag	3567
Db	3574	ACAGGTAAATCTCCAAAGTGAATAATTAAGAAAGTTAAAGAAAGCTTTAAATCTTACGA	3633
Oy	3568	aagttcctgccttgaagaaacagatctgcacactgctgtgc-----	3608
Db	3634	AATTTTCTCCAGAAAGCAAAAGTTTACAAACGTTGTATACCTCCACCTCAACAGATGTAAT	3693
Oy	3609	-----tgaactctgcctccagacaacactggagcgagctctcaactgaaga-----	3650
Db	3694	CCATCTCCATTAATCTGTAAAGGTAAGTGGTAGTTCAGGATCCACAAAGAAAGAACACAA	3733
Oy	3651	--gcctgcatactacatctgctgcagccgagctccaatacatatccacatctcagaagctc	3708
Db	3754	ATACCACTTACGAGCTCTTTATTAAAGATTTACAAACAGTATGATCAATCAACAAATATAT	3813
Oy	3709	gaagataagctgatactgacgcgtacatctgtgcatactctgcgcgagagcgaggagactac	3768
Db	3814	GACCAAGAGATGTATTCCTTAGTTGTAATACCATTTTGGAGAAATCCGAAATATATGAC	3873
Oy	3769	gaigacctgcgcgaagtgatcacccgttgagcgctgaacccttcctgtaattgataaact	3828
Db	3874	GAATATTTAGTTCAGTAGTAGTACCTGGAGAAGC---AATATCTGTCAATGGATATATATC	3930
Oy	3829	ctgctcaaaatcgsaacgatacgaagtgctctatctgaacactctgcaagcgctcat	3888
Db	3931	CTCTCAGGATTTGAAATATGATATGATGTTATATTATTTAAACCTTTAGCTGAGATATAT	3990
Oy	3869	aggtctctcaagaaacagctgsgaabaacgtaagactccaatgltcaacgttgaaaggac	3948
Db	3991	AGAAAGCTTAAAAAACAATTTGAAAAAACAATATTACATTATTAATTTAAATTTGACAGAT	4050
Oy	3949	attctgaacgcgctcttaataagagaaattcaagaacgctctggagagcgactg	4008
Db	4051	ATCTTAAATTCACGCTCTTAAAGAAACGAAATATTTCTTAGATGTATTTGAAATCTGATTTA	4110
Oy	4009	attccctataaagccgacgtactcctcttaactacgtctgtaagaccatacaagtctctc	4068
Db	4111	ATGCAATTTAAACATATATCTCTCAATGATATCATTTATTTGAAGATTTCAATTAATATTG	4170
Oy	4069	aataaagagaagaggataaattctgctagltacactatatacaagactccatcgac	4128
Db	4171	AATTCACAAACAAAAACACACTTTTAAAAAGCTTCAAAATATATTAAGAAATCACTAGAA	4230
Oy	4129	accgatatcaattctgcctaaatgctgtgcggtgatatacaagatccctgaagcgaaaaatac	4188
Db	4231	AATGATATTTAAATTTGCACAGGAAGGTATTAAGTTATTAAGAAAAGCTTTTACCGAAATAT	4290
Oy	4189	aagctgaacctgactctataaagaatatatacaagataag-----	4230
Db	4291	AAGATATTTAGATCAATTAATTTAAAAAAGTTATCAAGAAAGAAAGAGAGTTCCCATCATCA	4350
Oy	4231	-----caagcgagatgaataaataat	4251
Db	4351	CCACCAACAACACCTCCGTCACGACGAAAAACAGACGAACAAAGAAAGAAAGTAAGTTC	4410
Oy	4252	ctgcacctcttgataatacatctgaaaccctgtacaagacagtgaaacgacaaatcgactc	4311
Db	4411	CTTCCATTTTAAACAAACATTGAGACCTTATACAAATACCTTGTTAAATTAATTTGACGAT	4470
Oy	4312	cttgtaattcaaccggagggacgaagcgtcccaactactactaagagaagaagcaatgtgaa	4371
Db	4471	TACTTAAATTAACCTTAAAGGCAAAAGATTAAACGATTTGAATGTTGAAAAAGATGAACACAT	4530
Oy	4372	gttataatcaagagctgtaactaccataaacaatcaagaagcctgacgattccaag	4431
Db	4531	GTTAAATTAACCTTAACTTAACTTGAATTTAAAGCATTTGATGACAAATATGATCTTTTAA	4590
Oy	4432	aaaaatacaattctgctggaattctgacagcgtgtctaccgattataaccacaaacatctc	4491

Dd	4591	AACACTAACGACTTGTGGAAAGCATTTAAAAAATGTGTAATAATGATGATACGAAGAAAAGAATATG	4650
Oy	4492	ctgacaagatttcgtgccactggcatcgtgtgttcgaaaacccgcgaacaaagtgctgagc	4551
Dd	4651	CTTGCGCAAAATTACTTAGTACAGATTTAG--TTCAAATAATTTCTTAATPACAATAATATCA	4707
Oy	4552	aatcgctcgagggaacctgcaaggcacctgcgaacatctcccagcaaccaatgcgtgaag	4611
Dd	4708	AAATTAAATTGAAGAAATTTCCAAATATATGTTAAACATTTTCACAACACCATAATGCCTAAAA	4767
Oy	4612	aaacagtccccccagaataagcgcgtcttccagcatctggacagcgcgcaagagtgcaag	4671
Dd	4768	AAACAGATGCCGAGAAATTTCTGATGTTTCACACATTTAGATGAAGAAGAACAAATGTAA	4827
Oy	4672	tgtcttcctgaactaacaaaagaagagataaigtcgttgagagaaaccctaactgc	4731
Dd	4828	TGTTTTATTAAATTTACAAACAGANAGSGTATTAATGTGTGAATAATCCAAATCCTACTGT	4887
Oy	4732	aatgaacaatatggcgggtgtgagccgatctgaataatgcaccgagagaagacagcgctct	4791
Dd	4888	AACGAAATTAAGTGGATGTGATTCAGATGCGCAATGACGAAGAAGATTCAGGTAGC	4947
Oy	4792	aacgaaagaatacacatcgagtcgagtgtaactaagccgactctcatcactctcgacgg	4851
Dd	4948	AACGAAAGAAATACATGTCATGATGATTAACCTGATTTCTTAATCCACTTTCCATGCT	5007
Oy	4852	atttttgtccagctcaattccctggcatctctctctctcgtcgtacatcctcagtcgalc	4911
Dd	5008	ATTTCCTGAGTTCCTCAACTTCTTAGGAATATCATCTTATTAATCACTCATGTAAATA	5067
Oy	4912	ctgtacagctcatccta 4929	
Dd	5068	TTATACAGTTTCATTTAA 5085	
RESULT 12			
PEP190G1			
LOCUS	PEP190G1	5392 bp	DNA
DEFINITION	Plasmodium falciparum p190 gene for surface antigen precursor.		INV 22-JAN-1999
ACCESSION	X05624		
VERSION	X05624.1	GI:4186080	
KEYWORDS	antigen; glycoprotein; p190 gene; signal peptide; surface antigen.		
SOURCE	malaria parasite P. falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 5392)		
TITLE	Tanabe,K., Mackay,M., Goman,M. and Scaife,J.G.		
JOURNAL	Allergic dimorphism in a surface antigen gene of the malaria		
MEDLINE	parasite Plasmodium falciparum		
REFERENCE	J. Mol. Biol. 195 (2), 273-287 (1987)		
AUTHORS	2 (bases 1 to 5392)		
TITLE	Tanabe,K.		
JOURNAL	Direct Submission		
COMMENT	Submitted (21-Jan-1999) K. Tanabe, Osaka Institute of Technology,		
REMARKS	Biology, 5-16-1 Ohmiya, Asahi-Ku, Osaka, 535-8585, JAPAN		
FEATURES	Sequence revised by author		
	On Jan 26, 1999 this sequence version replaced gi:1926.		
	Location/Qualifiers		
Source	1..5392		
	/organism="Plasmodium falciparum"		
	/strain="MAD20"		
	/db_xref="taxon:5833"		
	/clone_lib="(lambda)NM1149 genomic"		
	136..192		
	136..5241		
	/note="p190 precursor"		
	/codon_start=1		
	/protein_id="CAA29112.1"		
	/db_xref="GI:4186081"		
	/translation="MKIIIFCSLFITINTQCYTHESYELVKKLEALDAVLTVTS		
	LFOEKAVLNEGTSYAVVTSTSPSSSVYSGGSAVASVAASGSGSVASGSGNS		
	RRTNPDSNSSLNTKTYADLKHRPOANTLTFTKEJLYEFLDTLNHMLTLISKNDGFFX		
	sig_peptide		
CDS			

Db 1258 TTAAGAGAAAAAATAAAGTTGATGTACACCTAAATCAAGATCCTACGAATCT 1317
 Oy 1195 gaatacaggaacctaataatccaaatggtgagctacccctgtctataacagt 1254
 Db 1318 GTTCAAAATACCAAAAGTTCCTTATCCAAATGGATTTGTAATATCTTTACCTACTCATGTAT 1377
 Oy 1255 atcaacaacgctctc-----aagcgctcaataagcttcggtgacttgatacccc 1305
 Db 1378 ATTCAATATTCATTTAGCTGCATGATATGATAAAATTCATATGGTATTTAATGAACTCT 1437
 Oy 1306 ttcgattatacgaagaaccccttaagaatatactacaagacatacgaagaagaagttc 1365
 Db 1438 GATATCAAGAAAAAATTAATCAAAAAATTAATACAGATATACGAAAGAAAAATATATTC 1497
 Oy 1356 atcaacgaatacgaagaagatcaaaattgagaagaagaattgagagtgacagaaga 1425
 Db 1488 ATTTAATTAACATTTAAAAACAATTTGATTTAGAGAAAAAACAATTTATCACACAAAAAGA 1557
 Oy 1426 agttagaagacgcagcaaaagctlaaacgatalcaactaaagatlatgaagaagctgctg 1485
 Db 1558 CAAATTAATAAATTTACTTCAAGATATATGAAAAGTCAAAAAAGATTTATCAAGATTTACTT 1617
 Oy 1486 aacgagatctatgattccaattcaacaataacatcgaccgacccaacttcgagaanaatg 1545
 Db 1618 GAAAAATTTTATGAATGAAATTTATATATTTATTTGACAAAGATGCTGATGATTAATA 1677
 Oy 1546 atggaagaacgctactcttacaagaatgagaaactgacacaccataaactcttgatctc 1605
 Db 1678 TTCACTGCAGATATATACATATATATGTTGAAAACAAGATATATATTAATTAATTTTCATCC 1737
 Oy 1606 tatgaagaattcttaagcataactcttagaagctcaccaaaagctcttaagtalatacggagac 1665
 Db 1728 TCTATTAATTTCTGTATATATATGTTCAAAAAATTTAAAAAGCTCTTTCAATCTTGAGAGT 1797
 Oy 1666 tactctcgcggaacaatggtgtgagagaagaactaaagatlatcaagaatctcataagt 1725
 Db 1798 TATCTTTTAAAGAAAGAAATTTCTGAAAAAGATTTTAATCATTAATTAATCTTTGAAAACT 1857
 Oy 1726 aagatcgaagaacgagatcgagacgctgtgtgagaacatlaaagaagaatgagaagaagctg 1785
 Db 1858 GGCTCTCGAAGCTGATATATATTAATAACAGAGAAATTAAGAGATGTAAGAAAAACAATTT 1917
 Oy 1786 ttgagaagaagaattacaagaagacgaanaataaaccagatgagaagaatccctgaggtctcc 1845
 Db 1918 CTAGAAAAAATTTTAAAGCACTAACATTCAGCAAAATG---CTTCCTTAAGAGATCT 1974
 Oy 1846 gatatgttaagaatcccaagctcagaaggtgtctctcaatggaacaagaatgtaactcaag 1905
 Db 1975 GATATTTGTAATAATTTACAGGTACAAAAGTTTATTAATTTAAAAAATATACAGACTTAACA 2034
 Oy 1906 aagatcgaacatctctgaagaagctgagtttaaaacataataatataatgctgcgaatgt 1965
 Db 2035 AAGATATGAATTTATTTTAAAAAATGCACACATTAAGATAGATTTATTCATCTACCAAAATAT 2094
 Oy 1966 tataagcagaagaataaagcagaacacatactactccatcgtatactacaagaagaatagac 2025
 Db 2095 TATAAACCAAAAATTAACCAAGCATATTTATTTAATTTGATTTAAAAAAGACAGATAGT 2154
 Oy 2026 aaacggaagaatgattacgccaagaatcgagagcctgatacagaagaagaagaagaacatt 2085
 Db 2155 AAATTTAAAGAAATTTATACCAAAAGTAAAGACATGTTAAAGAAAGAACAACTGTCTTTA 2214
 Oy 2086 aaacatgagaagcagctc-----agataactcgcagcctctcacagaagaagaatlaaac 2139
 Db 2215 TCAAGTATTATTCACAACTTTTACTTCACAGAACGCAAAACAACTGAAGATGGGGGTCACTCC 2274
 Oy 2140 ggaacaggtaccaccaagcccggaacacagcggttcaagctctcgaagcgatagcgtg 2199
 Db 2275 ACACACACATTAATCCCAATTCAGAGAGAAACAGAAATACAGAGAAACGAAGTAATACAGAA 2334
 Oy 2200 caagctcaagcacaagaagcagaagcagcagcctccagtgccagtgcccgltcc----- 2255

Db 2335 GAACAGTAGGACACACAAACGGTAAACAATTAACATTAACACCAAAAGAAATTCAGCA 2394
 Oy 2256 -----aagagctcaagctcaagtgctctacacacacagcctcgtgataaagaagcgag 2310
 Db 2395 CCAAAAAGAGTAAAGTGTCTTAAATTTCAATATGAAACATTAAGATATGACAAATTCACAA 2454
 Oy 2311 aatgtcagaacactgactacacttgagaagcctctatgagttccctgaatacactacac 2370
 Db 2455 GCCTTGACAAAACAGTTATCTTAAGAAATTAAGATGAATTTTAACTAAATCATATATA 2514
 Oy 2371 tgcacacaatatatccctgctctcacaagcactatgaaacgaagaatctcttaacaagtac 2430
 Db 2515 TGTCAATTAATTAATTTTATTAATCAAACTCTAGTATGAGCAAAAATTAATTAAGAGTATAT 2574
 Oy 2431 aagataaccaagaagaagagagatgaactgctcctcttgatccactggaacgtgctgtc 2490
 Db 2575 AATCTTACTCCAGAGAGAGAAAAAGAAATTAATTAATATGATGATCCATTTATTTATTT 2634
 Oy 2491 aatatcagaacacatcccgcttactgattctatglttctatgctcgaactcaacatctctc 2550
 Db 2635 AATATTTCAAAATTAACATACCTGCTATGTATTCATTAATATGATATGATGACAAATGATTTA 2694
 Oy 2551 tctcaactgttcatgagatalatagaagaagatggtctgcaacctglataaactcaaa 2610
 Db 2695 CAACATCTCTTTTGAATTAATATCAAAAAGAAATGATTTATTAATTAATCACTAACTAAA 2754
 Oy 2611 gacaacgacaagaattgaagacctctgagag----- 2641
 Db 2755 GAGGAAATCACATCAAAAATTAATTAAGAGAGCAAAAACAAATTAACGAAATCATCT 2814
 Oy 2642 -----aagctagaagaagctctccactctgttaaaactctctctc 2681
 Db 2815 ACATCCAGTCTCGAATAATACACCCGTAATACTGCTCAATCCGACACACAGTAATTC 2874
 Oy 2682 cagctccatgcaaccacgctctctcaacactcaagaaca----- 2720
 Db 2875 CAAAACCAAAATCAAAATGCAATCTCTACCAATPACCCAAATATGATGCTATCATCT 2934
 Oy 2721 -----gcccggaagtgagcgctcaagcagcagcactctcaactgaccacacttaactca 2775
 Db 2935 GGTCTGCTGTATGTTGAAGAAAGTATGATGCTCTTAACAGTATTTGCTATTTAGTAACGT 2994
 Oy 2776 ctgaaactgtttgagaacatctctctctcgcgcaagaataagaacat---ctaccaaga 2832
 Db 2995 TTGAAGAGTATGTTAGTCTCTTAATCTTGAAATTAACATTAAGTACCTAATCATTA 3054
 Oy 2833 ctatttgacaagaatlgctcgagaactctcaggaagaatctcgaagaagcagcagaaca 2892
 Db 3055 ACCATTTCTACAAACAGAGATGGAATAATTTTATGAAATATTTTAAAAATATGATATCC 3114
 Oy 2893 ttctatacgaagaagcttcaactactctgtaactcttaagaacgcatgatalatacactctc 2952
 Db 3115 TATTTTAATGATGATATCAACAAATTCGTAATTAATTTCAAAAGTAAATTCACAGTTTG 3174
 Oy 2953 aacgattgaatctlaaacgtaagaagcttgaagaagacatcaatlaaagcttgaagaagacactg 3012
 Db 3175 ACC---GAACACCAAAAATTAATCATTAATGATGAATTTAAAAATTAAGATACTTTA 3231
 Oy 3013 caactgagctctgacactgttaacaagaatgaacactggaagcagcagcctctgcagaag 3072
 Db 3232 CAGTTATCATTTGATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAG 3291
 Oy 3073 aagaagacagctcgcaagtlataagatcgaagctgaagctgtaactcgtcacaagaagcag 3132
 Db 3292 AAAAAAGAACTTGCCCAAGACAAATTCGAATTTAAAAAATTTACTTTTAAAGAACAA 3351
 Oy 3133 ctgnaaagcaactcaactcaactgacaatccgaaacacgtaactgcaagactctcagtg 3192
 Db 3352 TTAGAATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3411
 Oy 3193 tctctcaacaagaagaagcagagatcgccgagacagagaagaacactctggaagaaccc 3252
 Db 3412 TTCTTTAACAAAAAAGAAAGCTGAATTAAGCAAACTGAAATACACATTTAGAAAAACACA 3471

[illegible]

OY	310	gcaagcgcggtlccgcggaaacagltgaagaacccaatccatctgcgaacctctagcgattcc	369
Db	403	GCTTCAGGTGGTTCAGGTAATTCGAAGACGTCACAAATCCCTTCAGATTAATTCAGGTAAATTC	462
OY	370	gaqccaaagtlctccgcgcacccccaagcaacggatgaagaactatctccctactcaag	429
Db	463	AATACATAAATTACGCTGATTTTAAACATAGAGGTTCAAAATTTACTGTCTCACTTTAA	522
OY	430	gaqctgaagtaaccaagltgtctgcacctactaatacatalcgtcgcacctgtgtataac	489
Db	523	GAACTCAAATATCCCGAAGCTTTGATTTTAAACCAATCATATGTTACCTTAATCAAAAAT	582
OY	490	atctatgcttcaaatatctgcattgcgcgtctacgaagagatcaatgaactccgtacag	549
Db	563	GTTGATGGTTTCAATATTTAATTTGATGATATGAGAAATTAATGCAATTAATATTTAA	642
OY	550	ltgaattctactcgcactgtctcgaagggccaactgaatgaagtttgcgcactgactat	609
Db	643	TTAAACCTTTATATGATTTATTAAGAGCCAAATTAATGATGATCATGTGCCAATGTAT	702
OY	610	tgtcaaatlccatcattcatttgaagatcagagcccaagagltgacogtaltgaagaagltg	669
Db	703	TGTCAAATACCTTTCAATCTTTAAATTCGCGAAATGAATTAAGCGTACTTTAAAAATTT	762
OY	670	gtcttcgatatcgcgaagcctctgcgaacatcaatgaacatgtctggaaaga.tggaagat	729
Db	763	GTTTCGGATATACAAAACCATTAAGCAATATTTAAAGATATATGTGAGAAAATGGAGAT	822
OY	730	tatatlaaaagaatlaagaagaccatcgcagagacataaacagagctgatlgaagaatccaa	789
Db	823	TACATTTAAAAAATTAACACCATTAAGCAAAATTAATGAATTTAATTTGAAGAAGTAAAG	882
OY	790	aagaccatlagacaanaataagaatctgaacaaagagaagaagaagaagatctgacag	849
Db	883	AAACCAATTGATCAAAATTAAGATGCGAGATTAATGAAAGAAAGAAAAAATTAATACCA	942
OY	850	gcccagtagcagcctgtccatctatacaacaagcttgaagaagcccatcaactcagc	909
Db	943	GCTCAATTAATCTTTTATTTATTTACAAATTAACAAATTTCAAGAAAGCACATTAATTAATTAAGC	10020
OY	910	gtactcgggaagcgcatagacacccctcaagaagaatgaaatatalcaagaagactgtctac	969
Db	1003	GTTTGTAGAAAAAGCTATTGACACTTTTAAAAAAAATGAAAACATTAAGAAATTACTTGA	10620
OY	970	aagattaaatgaatlaagaaatccctccgcca--gccaactcttgggaacaccccctaacag	10260
Db	1063	GATATAGTATAAATTAATAAACAGATGCCGAAAACCCACATCTGGAAGTAAACCAATCT	112220
OY	1027	ctctcgtgacaagaacaaagaatagagag--caagagaagaagatacaagaagctgcc	10830
Db	1123	CTCCCTGGAAATTAAGAAAAAAGAGTCGAGGACGAGCAAGAAAAAATAAAGAAATTTCC	11820
OY	1084	aaaacatlaagtlccaacatagatctctctcttaccgactccctctgaagcttgaagtactac	11430
Db	1183	AAAACATTAATAATTTAACTTGAATGATGATTTATTTACGATCCACTTGAATTAACATTAATAT	12420
OY	1144	ttgagagagaagaatlaagaatataagacatctccgccaagctga-----gacaaag	11940
Db	1243	TTAAAGAGAAAAAATTAATAAAAGTTGATGTAAACCTTAATTAACAAAGATCTACGAAATCT	13020
OY	1195	gaatcaaacggaactaatgatataatcccaatgggtgagtgacccctctgtctataaagat	12540
Db	1303	GTTCAAATACCAAAAGTTCCTTATCCCAATAGATTTGTATATCTTTTACCACTCACTAT	13620
OY	1255	atcaacaacgctctc-----aacagctcaatagatctcgtgtgactgtg.tlaaccc	13050
Db	1363	ATTCAATTAATCATTAAGCTGCGAGATTAATGATAAAATTTCAATATGGTGAATTAATGAATCT	142220
OY	1306	ttcgattatacgaagaagaccctctaaagaatatctaacacagcaatgagagaagaagatct	13650
Db	1423	GATACATAAAGAAAAAATTAATGAAAAAATTAATTAAGATTAATTAAGGAAAGAAAAATATTTTC	14820

Db	3637	AAATATCTTTCTTATFCAAGAGATTCATCATCTTTAATTACGAATTAATAAGAGTAATA	3696
OY	3433	aagaaacaagaactacacccgcgaatagcccaagcagaaataataacagaagtgaataagca	3555
Db	3697	AAAATAAATAAATTTATACAGGTAATTTCCAAAGTGAAATAATATAGAAAGTTTACAGACT	3756
OY	3553	ctggaactctcaagaagcttcgcctcgaagagaacagatgctgcacattgctgctc----	3608
Db	3757	TTAAATATCTTACGAAATTTTCTCCAGAAAGCAAGTTTACACAGTTTAACTCCACCT	3816
OY	3609	-----tgaaatctgctccgaacaacatcctggaagcagctca	3641
Db	3817	CAACAGATGTAACTCCATCTCCATTAATGTAAAGGTTAAGTGTAGTTCAAGATTCACA	3876
OY	3642	acctaaaga-----gcctgcatctactcatgctcgagccgagctccaaataaattacc	3693
Db	3877	AAAGAGAAACACAAATTAACCAACTTCAGGCTCTTTATTAACGAAATTAACACAGCTAGTA	3938
OY	3694	acattctaaagcgcgcgcagcttgagctgaaatgaaatgataatgctgctactcttgaggag	3755
Db	3937	CAATTACAAATTTATGACGAGAAAGATGATCTCTTGTTGTTATTAACCATTTTGTGAGTA	3996
OY	3754	aagcgaaggaactcagaatgaactcgccgaagtgatgacacggtgaggtctgcaactctcc	3813
Db	3997	TCCGAGATATATGACGAATATTATGATCAAGTATGTAATGATTAACCATTTTGTGAGTA	4055
OY	3814	gtgattgataacatctgctccaaaatcggaaacgaatacgaagtgtctatactgaacct	3877
Db	4054	ACAAATGATTAATATCTCTCAGGATTTGAAATAAGAAATATGATTAATATTATTAACCT	4113
OY	3874	ctgcaagcgcgtctaagctctcctaagaacagctggaagaaatagctgtagccttcaat	3933
Db	4114	TTAGCTGAGATATATACAGCTTTAAAAAACCAATTGAAAAAACCTTTTTCATTATAT	4173
OY	3934	gtcaacgtlgaaagacatctcgaacagcgccttaataagaagagaaatttcaagaagctc	3993
Db	4174	TTAATTTGAACGATATCTTAATTAATTCACGCTTAAGAAAGCAAAATATTTCTTAGATGA	4233
OY	3994	cttgagagcgcgactgattccctataaagaacctgcactcctctaactagctgtgcaagac	4053
Db	4234	TTAGATCTGATTTTAATGCAATTTAAACATTAATCTCAATTAAGAAATCACTATTTCAGAT	4293
OY	4054	ccatacaagctccccaataagaagagagaggaataaattctgctgaattgaacactatc	4113
Db	4294	TCATTTTAATTAATTATTCAGAACAAAAAACACTTTTAAAAAGTTACCAATATATA	4353
OY	4114	aaggaactcaccgaacacagatataatcttcgcaatgagtgtcgtggagttatacaagctc	4173
Db	4354	AAAGATCAGTAGAATAATGATTAATTAATTGCAACAGAAAGATATAGTTATTATGAAGA	4413
OY	4174	ctgagcgaanaatacaagctcgaacctgactctataataaagbataatcaacgaataag---	4230
Db	4414	GTTTATGCGAATTTAAGGATGATTTAGATTCATTTAAAAAGTTATCAAAAGMAAAG	4473
OY	4231	-----caaa	4233
Db	4474	GAGAAATTCATATCAACACAAACACACTCGCTCACGCAAAAAACAGACGAACA	4533
OY	4234	ggcgaagaatgaaaaatactgccccttcctgaataacatgaacgaacctgttgaagaacgtg	4293
Db	4534	AAGAAGGAATAAGTTCCTTCCTTTTAAACAACAATTGAGACCTTATACAAATACCTTA	4593
OY	4294	aacgaacaatcgaacctcttcgaatcaatcactcgtgagccgaagctcctcaactatacttac	4353
Db	4594	GTTATATAAATTTGACGATTTACTTAATTAACCTTAAGCAAAAGTATAGCATTTGTAAGTT	4655
OY	4354	gagaagagcaatgtggaagttaaaatacaagagcgtgaactaccctaaacaatccaagac	4413
Db	4654	GAAAAAGATGAGCAACATGTTAAATTAACCTTAAGGATTTAAAAACAATTGATGAC	4713
OY	4414	aagcgtgcaagattctcaagaanaaataaataattctgctgaattgcaagctgtctacgat	4473
Db	4714	AAATATGATCTTTTAAAAACCTTAACGATCTGAGACATTTAAAAATTTGTAATATAT	4773

OY	4474	tataccacaacaaatccctcgaccagaatttcgtccacgagcatgytgcctogaaaactc	4553
OY	4474	GATACGAAAAAAGTATTGCCTGGCAAAATTACTTAGTACAGGATTG---TTCCAAATTTT	4830
OY	4534	gccaaaaacagtgcgaaccaatcgtccgcgaacacctgcaggcatgctgaacatccc	4593
Db	4831	CCTTATTCATTAATATTCAAAATTTAATTGGAAGAATAATTCAGATATGTTAACATTTCA	4890
OY	4554	cagcaccaatgctlgaagaacacagtgccccagaatagcgctgttttaagcatctggac	4653
Db	4891	CAACACAATGCCGTAAAAAACAATGTCCAGAAAATTCGATGTCTTGACGACATTTAAT	4950
OY	4654	gaggcgaagaagtcgaagtgctccctccaactcaaacgaagaagagataaagtcgtagg	4713
Db	4951	GAAGAAGAGAGATGTAAATCTTTATTAAATTACAAACAGAAAGTGATTAATGTGTTGAA	5010
OY	4714	aaccacaacctacctcgaatgaanaacaaatggcgggtgfygaacgcgaatgaatgcac	4773
Db	5011	AATCCAATCCTACTTGTATACGAATAATATGCGATGTGATGACAGATGCCACATGTACC	5070
OY	4774	gaggagacacagcggctctaacgysaaagaataatcacatgcgaatgctactaagcccgactc	4833
Db	5071	GAAGAAAGATTGAGGTACACGACAGAAAAGAAATCACATGTGAATGTACTAACCTGATCT	5130
OY	4834	tatcacctctgcgacggagattttctgcacgctcaacttcctctggacatccctctg	4893
Db	5131	TATCCTACTTTTCAGTGTGATTTTCTGTCAGTGTCTCTAACTTTTGAAGATATCATCTTA	5190
OY	4894	ctgatccctatgctgataccctgtacagctcatcataatagatcatg	4939
Db	5191	TTAATCTCAATCTTAAATATTATACCTTTTCATTTAAAAAATGTAGG	5236
RESULT 14			
PMEZSAIB			
LOCUS	5312 bp	DNA	INV 10-MAR-2001
DEFINITION	P.falciparum gp190 (MSAI, MSP1, PMMSA) for precursor of major merozoite surface antigens.		
ACCESSION	Z35327 U09209		
VERSION	Z35327.1 GI:929795		
KEYWORDS	gp190 gene; merozoite surface antigen; MSAI gene; MSP1 gene; PMMSA gene.		
SOURCE	malaria parasite P. falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	1 (bases 1 to 5312)		
AUTHORS	Pan,W., Tolle,R. and Bujard,H.		
TITLE	A direct and rapid sequencing strategy for the Plasmodium falciparum antigen gene gp190(MSAI Mol. Biochem. Parasitol. 73 (1-2), 241-244 (1995)		
JOURNAL	96123395		
MEDLINE	2 (bases 1 to 5312)		
REFERENCE	Tolle,R., Bujard,H. and Cooper,J.A. Plasmodium falciparum: variations within the C-terminal region of merozoite surface antigen-1 Exp. Parasitol. 81 (1), 47-54 (1995)		
AUTHORS	95354793		
TITLE	3 (bases 1 to 5312)		
JOURNAL	Tolle,R. Direct Submission Submitted (18-JUL-1994) Ralf Tolle, Abt. Prof. H. Bujard, Zentrum für Molekulare Biologie, (ZMBH), Univ. Heidelberg, Im Neuenheimer Feld 282, Heidelberg, 69120, Germany On Aug 5, 1995 this sequence version replaced gi:535249.		
COMMENT	Location/Qualifiers		
FEATURES	1..5312		
SOURCE	/organism="Plasmodium falciparum"		
	/isolate="NF54"		
	/db_xref="taxon:5833"		
	/chromosome="9"		
	/clone="3D7A"		
	/dev_stage="blood stage"		

sig_peptide 109..165
/gene="gp190, MSA1, MSP1, PMMSA"
gene 109..5271
/gene="gp190, MSA1, MSP1, PMMSA"
CDS 109..5271
/gene="gp190, MSA1, MSP1, PMMSA"
/product="precursor of the major merozoite surface
antigens"
/protein_id="CA84556.1"
/db_xref="GI:929796"
/translation="MKIIIFLCSEFLFIINTQCTHESYOELVKKLEALDELAVLGY
LFOEKMYLNEEITLTGASASGASGASGASGASGASGASGASGASGASGASG
SGSGTSSPSRNLTPRSNTSGASGASGASGASGASGASGASGASGASGASG
ELFLTNHMLTLCNINHGFKLIDGEEINELKLNLYFELRLAKLMDVCANDCOI
PFLNLRANDELVLKLVFGYRKPDLNKTGVGMEDIKRNKTITANIINLEIGSK
TIDONKNANDEGKRLQAOYDLSTYKQLEAHNLISVLEKRIIDLKKNENIKL
DKINEIKNPSPANGTNPDLDRNKKEEHEEKIKELAKTIKENIDLPDPLEY
YLREKKNKVDYTPKSODPTKSVQIAPKVPYPGIYPLPLDTHNSLADNDKNSGDL
MNPHTKEIKETITDNKERKIFINNIKKIDLEKKNHTKEDONKILDEYKSKD
YELLERFTEKERNNEDKDYDKIFARITYNWKORNNKFSNSNVYNQVLK
ALSTLEDSLRKGISEKDFNHVYLTGLEDIKLKEIKSSENKILEKPKGLTHS
ANGSLVSDIVKLOVOKVLLIKKIEDLKLFLKNQKLSIHVPNIKPNKPEPY
YLIVLKEVDLKEFIPKVKMLKEQAVLSITOPVLASETTEDEGSHSTLSOG
ETVEETEETEETVGHHTVTITLPPSPKPKVEMSEIKHSNDNSOALTRION
YLKIDELFPTKSYICHKYIIVSNSSMOKLEFVNLPEEENELKSCPDLLTRION
NIPAMYSLYDSMNDLOHLEFELYOKEMITYLHKIKKENHILKLEBKQITGHSST
SPGNTVNTAOSAKHSNSNOONASSTNTONGAVSSGPVAVESHDPVLSISND
LKGIVSLNLTAKTSKPNPLITSTTEEMKPEENILKNNDYENDIKOPVNSVIT
GLETOKNALDEIKLKLIDLSFDLYKRLKLDLKNKKEGIDOKMOIKLTL
KEOLESKLNSLNPHNVLONFVFNKKKEIKETENTENTILKHYGLVKNY
GESSPLTISSEVSIOTEDNVANLEKPLSLIDKLNHLGKRLKSLSSGHLIL
TELEKYEKNKNTGNSPSENKKYNEALKSTENLPKAVTTPVPPDPVPSLAV
RVSSGSGSTKEQIPTSGLTLEQOVYQONDEEDDSLVLPITRESDNDEYLD
QVNTGEAIVTMDNLSGFENEVDYIYLPAGVRSRKQIENINIFENLNDILN
SRLKRAYFLDVLSDLMQFKNHISNEYIIESFKLNSSEOKNLLKSKYKIKESVEN
DIKFAOEGISYEVKLVAKYKDLDESIRKVIKEEKPESSPTPPSPKADDEKES
KELPFLNIEFLVNLVKNIDYLLINKAKINDONVEKEDVAHVITKISDLKIDKI
DLFNRPYDEAIKILINDTRKMDGLKLSGLVONPNTIISKILBEKFEEDMLNISQ
HOQYKQCPENSGCFRHLDEREBCKLINTVROGEDCENENPNCNENNGKQDADATC
TEEDSGSRKKITECTKPDSPJEDGJFCSSNMLGISFLILMLILYSP1"

BASE COUNT 2328 a 703 c 705 g 1576 t
ORIGIN

Query Match 25.8%, Score 1272.2, DB 96, Length 5312;
Best Local Similarity 55.9%, Pred. No. 7.2e-261;
Matches 2900; Conservative 0; Mismatches 2018; Indels 267; Gaps 15;

Qy 10 atgaataatcttctctcgtcattctcgttttatcatcaataactcagcgctg 69
Db 109 ATGAAGATCATATCTCTTTTATGTCATTTCTTTTATATATAACAAGTGTGA 168
Qy 70 acccaacatccatcagagcttgcttaagaacatcgaagcttggagaatccgctct 129
Db 169 ACACATGGAAGTATACAGAACTTGCTCAAAAACCTGAAGAGTATGACATATG 228
Qy 130 accggaatacagcgtgtccagaagagaagatgctgtaataga----- 174
Db 229 ACAGGTATATAGTTTATTTCAAAAGGAAAAAATGATTAATGAGAGAAATTACTACA 288
Qy 175 -----ggagacgagtggacagcgcggttacaaccgc 204
Db 269 AAAGTGCAAGTGTCAAAAGTGTGCAAGTGTCAAGTGTCAAGTGTCAAGTGT 348
Qy 205 aacacgggttctlaaaggcttgctgtagcagtggtccggtggtctgctgctcctg 264
Db 349 GCAAGTGTCAAAAGTGTGCAAGTGTGCTCAAAAGTGTGCAAAAGTGTGCAAGT 408
Qy 265 ggttcgctgcctccgagcgagctggtgcatcaggtgctcagtgcaagcgcggttcc 324
Db 409 GGTCCAAAGTGTGCAAGTGTGCAAGTGTGCTGTTCAAAACATTAACTGCTGTCA 468

Qy 325 gggaaacagtcgaagaaccatccatctgacaactctgagatctccagcgcaagtcctac 384
Db 469 AATFACTTCTACTGCTGTCAGCAAGCCCTCCAGCTGTGCAAGCGATTCAGATGCTTAATCTTAC 528
Qy 385 gccgacatcgaacacagagtgagaactatctctcactatcaagaagctggaagcca 444
Db 529 GCTGATTTAAACACAGAGTACGAATTAATCTGTTACATTAAGAACTCAAAATATCCC 588
Qy 445 cagtggttcgacctactaataatcgtctgacactggtgtgataacatcgaagcttcaa 504
Db 589 GAACCTTTGATTTAAACCATATATGTTAACTTTGTGTGATAAATATCATGCTTTCAA 648
Qy 505 tatcgtatgagcgttgcagaagatcaatgaaccgcgtgaagaagttgaattcctactc 564
Db 649 TATTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 708
Qy 565 gacttgctaaaggccaaactgatacgtcttgccaaatgactatgccaatctccatc 624
Db 709 GATTTATTAAGCAAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 768
Qy 625 aattggaagatcagaagccaagagltggacgtatctgaagaagltggtctcgatatacgc 684
Db 769 AATCTTAAATTCGTGCAAAATGAAATGAACTTAAATAAACTGTGTCGATATAGA 828
Qy 685 aagcctctcgaacaatcagaagacaatlgggaaagatgggaagatattataaagaag 744
Db 829 AAACCATTTAGACAAATTTAAAGATATGATGAGAAATGGAAGATTTACTTTAAATAAAT 888
Qy 745 aagaagaccatcgagaacatlaacagctgtagcagaagatccaaagaagccaataga 804
Db 889 AAACACACATGACCAATTAATTAATGAATTAATTTGAAGAGATGAACAAATGATGATCA 948
Qy 805 aataagaatgcaaccaaaggaggaagaagaagaagctgtacagggccagtagaacctg 864
Db 949 AATAAGAAATGCGATTAATGAAAGAGGAGGAAAAAATTAATACCAAGCTCAATATGATCTT 1008
Qy 865 tccatctataaagaagcgttgaagaagccatccatcgaagctgtagaagaagcgc 924
Db 1009 TCTATTTACAAATTAACATTTAGACAGACACATTAATTAATGAGGCTTTTAAATAAATCT 1068
Qy 925 atagacacccctcaagaagaatgaataatcaagaagacgtctgcgaagaatlaatgaatc 984
Db 1069 ATTGACACTTTAAAAAATAATGAATAATGAATAATGATGATGATGATGATGATGATGAT 1128
Qy 985 aagaatcctcgcgcagccaactcgtggaaacacccctcaacagcgtctggaagaagaag 1044
Db 1129 AAAAATCCCCACCGCCGCAATTCGTGAAATACCAAAATACTCTCTGATGAAGAACAA 1188
Qy 1045 aagatagaagagcaagagaagaagatcaagaagatccgaagaacatlaagtlcaacata 1104
Db 1189 AAAATCGAGGAACGAGAAATAAATAAAGAAATTTGCCAAATCATTTAAATTTAACATT 1248
Qy 1105 gattctctcttactgatacccttgaagctgtagtactactcttgaagaagaagaatla 1164
Db 1249 GATAGTTTATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1308
Qy 1165 atagacatctccgcgaagctga-----gacaaagaaatcaacggaactaagaa 1215
Db 1309 GTTGATGTGTAACCTTAATATCAACAGATCTTCAAGAAATGTTCAATACCAAAATCTTCT 1368
Qy 1216 tatccaatgggtgtagtactccttgccttacttaacgatatcaaaagcctc----- 1268
Db 1369 TATCCAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1428
Qy 1269 --caacagcctcaatagctctcggtgacttgatlaaacccctcgtatcaacgaagaagcc 1326
Db 1429 GATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1488
Qy 1327 tctaaagatactacacagacaatgaagaagaagattatcaacgaatacaagaagaag 1386
Db 1489 GAAAAAATTTATTAACAGATTAATGAAGAAAGAAAAAATTCATTAATTAATTAATTAATTA 1548

OY	1387	atbaaatcttggaagaagaanaattggagtgagcagaagaagaagttaogaagaacgcgcgcga	1446
OY	1388	atbaaatcttggaagaagaanaattggagtgagcagaagaagaagttaogaagaacgcgcgcga	1446
Db	1549	ATTGATTGAGAGAAAAAACATTAACTACACAAGAAAGCAAAATATMAAAATTTACTTGAA	1608
OY	1447	agctcaaacgcatcactctaaagagatgaanaagctgtgcaagagatcctatctccaa	1506
Db	1609	GATTATGAAAGTCAAAAAGAGATATGAGAAATTTCTTGAAAAATTTTATGAAATTGAA	1668
OY	1507	ttcaacaataacatcgaccctgacccaacttcgagaaaaatgatvggaaaaacgctactctac	1566
Db	1669	TTTATATATATTTTGTGACAAAGATGTCGTGATGATAAATTTTCAGTGCAGATATCATAT	1728
OY	1567	aaagvggagaacatcgacacaccctaactcttgcatctctatgagaattcttaagataat	1626
Db	1729	AATGTTGAAAAACAAAGAT	1788
OY	1627	cttgaagaagctcaccaagaagctctaagatatatgagactatctctcgcggaacttg	1686
Db	1789	GTTCAAAATTTAAAAAAGGCTCTTTCTATCTTTGAAGATTTATTTCTTTAAAGAAAGAA	1848
OY	1687	gtvggaaagaactaaagatctaaagaatctcaagaatctcaagaatcgaaaaagatcgag	1746
Db	1849	TCTGAAAAAGATTTTATTCATTTATTTATTTACTTTGAAAGCTGCCTGAAACCTGATAT	1908
OY	1747	acgcctgtctggaacatctaaagaagatgaaagacagctgtcttggagaagaagatcaaa	1806
Db	1909	AAATTTAACGAGAAATTAAGAGATGAGAAACAAAATTTCTGAAAAAAATTTTAAAG	1968
OY	1807	gacgaaataaacacagatctggaagaatcctcgagagctctcgatattgttaagctcaag	1866
Db	1969	CTAACACCTTCAGCAATG---GTCTTGAGAGTCTCGATATTGTAAATTTACAAATA	2025
OY	1867	cagaagctgctcctcatgaaacagatgtgatgaaatcgaagaagactcaactcatctgag	1926
Db	2026	CAAAAAGTTTATTAATTTAAAAAATVGAAGACTTAAGAAGATGAAATTTATTTTAA	2085
OY	1927	aacgtggaagttaaaacataatatacatgtgcccgaatagttataagcagagagaataag	1986
Db	2086	AATGCACACAACTAAAAGATGATGATTTCAATGCCAAATATTTTAAACCCAAATTAAC	2145
OY	1987	gaaccaatacaccctcatcgactcaagaagaagatagcaaaactigaagtgtaagctcc	2046
Db	2146	GACCATATTTTAAATTTGATTTAAAAAAGAGTGTGATTAATTTAAAGAAATTTTACCA	2205
OY	2047	aaagctcgagagctgcatcaacggaagaagaagaacattaaacctgaagcagcgc---	2102
Db	2206	AAAGTAAAAAGCATGTAAAGAAAGAAACAACGCTGTATCAAGATATTAACAACACTTTA	2265
OY	2103	--agataactcggagccttccacagaagaagataaacggaagagctaccacaagccc	2160
Db	2266	GTTTCACGACAAACGGAAMACMTGGAAGATGGGGCTCACCTCCACACACATTTATCCAA	2325
OY	2161	ggaacaacagagccggtctcagctctcgaagcgatagagctgcaagctcaagcacaagaag	2220
Db	2326	GGAGAAACAGAAATGACGAAAGAAACAGAAAGAAACGAAAGAAACGTGAGACACACACA	2385
OY	2221	aagcagagc-----acagcctcagatgctgcagtgccccgttccagagagcttaagctaa	2271
Db	2386	ACGGTAAACATTAACATTACATCCACCAACAACATCAACCAAAAAGAAAGTAAAGTTG	2445
OY	2272	gtgcctacaaccaagcagctcctgtgaaataacaagaaccggaatgtcagcaaaactgga	2331
Db	2446	GAAATTTCAATAGAACATTAAGAGTAAATGCAAAATTCACAAAGCTTGACAAAACAG	2505
OY	2332	cttgaagaagctctatgagcttcctggaataactctacatctgtcacaanaatatatctgc	2391
Db	2506	CTAAAGAAATTTAGTGAATTTTAACTTAATCATATATATATATATATATATATATAT	2565
OY	2392	tctcaacagcatctgaacggaagaatctcttaacagtaacaagataacaagaagaagag	2451
Db	2566	TCAAACTCTGTATGAGGACCAAAAATTTATTAAGGTATATATATCTTACTCCAGAAAGAA	2625
OY	2452	agtaaacgtctcctctgtgataccaatcgagcctgctgttcaatatccagaacaaactcc	2511

[illegible]

D	3703	GGATTACATCACTTTAATTACTGAAATTAAAGAAAGTAAATAAAATAATTAATATACAGT	3762
O	3514	aatagcccaagcgcgagaaataacagacgtaagaacacacgtgaactctacaagaagtc	3573
D	3763	AATTCTCCAGTGGAAATAATATAAGAAAGTTAAAGCAAGCTTTAAATCTTACGAAATTTT	3822
O	3574	ctgcctgaaggaagaagctgtgcgcactgtgtgtc-----	3608
D	3823	CTCCGAGAAAGCAAAAGTTTACACAGTTGTAACTCCACCTCAACAGATGTAACTCCATCT	3882
O	3609	-----Tgaatctgtgtccgcagacactgtgagcagcttaactaaagaag-----ct	3654
D	3883	CCATTATCTGTAAAGGTAAGTGGTAGTTCCAGGTCCACAAAAGAAACAAACAAATACCA	3942
O	3655	gcactactcatgtctcgagcgcgagtcacaatacatattacacactctcagaacgttcagat	3714
D	3943	ACTTCAGGCTCTTATTATTAACGAAATTACAAACAGTAGTGATCAATTTACAAAATTATGACGA	4002
O	3715	gaagtcgatgtacgtcatcatctgtgcctatctctgcgcgagacgcgagagagactaagatgac	3774
D	4003	GAAAGATGATTCCTTAGTTGTAATTAACCATTTTGGGAAATCCGAAGATATATACGCAATAT	4062
O	3775	ctgcgcgaagtggtgcacgcggtgaagcgtgcataccctctcgttgattgaataagctgtgcc	3834
D	4063	TTAGATCAAGTAGTAAGTACTGGAGAACG---AATATCTGTACATAGATTAATTCCTCTCA	4119
O	3835	aaatacgagaacgaatacgaagtcctatctgaacacctgtgcagcgcgtatagact	3894
D	4120	GGATTTGAAATTAAGTAATGTATGATTAATTTAAACCTTTAGCTGAGATATATFAGAACG	4179
O	3895	ctcaagaacacgcctgcggagataacgtgatagaccttaatgtcaacgttgaagacactgtg	3954
D	4180	TTTAAAAAACCAATTTGAAAAAAACATTTTTACCTTTAATTTAAATTTGACACATCTCTTA	4239
O	3955	aacgcgcacttaataagagagaagaattctaaagaacgctcttgagagcagacttgatctcc	4014
D	4240	AATTACAGCTCTTAAGAAACCAAAATATTTCTTGATGATGTATTGAATCTGATTTAATGCA	4299
O	4015	tataagacactgcctctcttaactacgttgcataagagaccataaagtlctccataataa	4074
D	4300	TTTTAAACATATATCTCTCAATTAAGTAATGATATTTGAAGATTCATTTAAATTAATGAAATCA	4359
O	4075	gagagaagggataaattctgtctgagttaacatatagaagactccatgcagaacgat	4134
D	4360	GAAACAAAAAACACACTTTTAAAAAGCTTCAAAATATATTAAGAAATCAGTAGAAAAATGAT	4419
O	4135	atcaatttgcctaaatgtgtgcgtgggatttcaagaatccctgcaggaanaataacaaagct	4194
D	4420	ATTTAATTTGACGAGGAGGTATTAAGTTATTAAGAAAGGTTTTAGCGAAATATTAAGAT	4479
O	4195	gacctgacctatataaagaatatactcaagataag-----	4230
D	4480	GATTTAGATCAATTTAAAAAGTTATTCAAAGAAAGAAAGACAGAACTCCCATCATCACCA	4539
O	4231	-----caagcgcgaatgaaataatactg	4254
D	4540	CCAAACAACACTCCGTCACCAAGCAAAAACAGACGAAACAAAGAAAGTAAGTTGCTCTT	4599
O	4255	ccctctctgataataatcgaagaacccgttacaagacagtgaaagacaaaatgcactcttc	4314
D	4600	CCATTTTAAACAAACATTTGAGACCTTATACAAATTAACCTTGTTAAATTAATTTGACACATTAAC	4659
O	4315	gtaataccactgcggagcgaagtcctccaaactaacttaacgaagaagcaatgttgaaagt	4374
D	4660	TTAATTAACCTTAAGCAAGAAGATTAAACGATGTGAATGTTGAAAAAGATGAGACACATGTT	4719
O	4375	aaaaatcaagagactgaaactacctcaaaaacatccaagaacgctgtgcagattcaagaa	4434
D	4720	AAAATTAACCTTAACCTTATGATTTTAAAAAGCAATTGATGACAAATATGATCTTTTAAAAAC	4779
O	4435	aatacaatttctgcgaatattgagacacctgtctaccgatataaaccaacaactctctg	4494
D	4780	CTTTACGACTTTCGAAGCAATTTAAAAAATTGATTAATGATGATGACAAAAAAGATATGCTT	4839

OY	4435	accaaagtctcgctccactgacatgcatggtctcgaanaacctgcgcaaaacagtgctgaagcaat	4555
Db	4840	GGCAAAATTACTTACTACAGAGTTGG--TTCAAAATTTCTTAATACAAATAATATCAAAA	4896
OY	4555	ctgcctgcacgcgaaccctgcagggcatgctgaacatctccagcaaccaatgcg tgaagaa	4614
Db	4897	TTAATTGAGGAAAAATTCCAAAGATATGTTAAACATTTCACCAACACCAATGCGTAAAAAA	4956
OY	4615	cagtgccccagaaatagcggcctgtttcagcatctggacagcgcgcgaagatgcaagtgc	4674
Db	4957	CAATGTCGAGAAATTTCTGGATGTTTCAGACATTTAGATGAAAGAGAAATGTAAATGT	5016
OY	4675	ctccctgaactcacaagaagagatgaaatgctgtagagaaaccaaccatccctgcgat	4734
Db	5017	TTATTTAAATTACAAACAAGAGGATGATTAATGTGTGAAATATCCAAATCTACTTGTAAC	5078
OY	4735	gaaacaatgycgggtgtgaacgcgatgtcctaaatgcaccgaggaagacagcgcctctaac	4794
Db	5077	GAAATTAATGTGTGATGTGATGCGAGATGCCACATGTACGAGAAAGAAAGATTCACAGTGTGACAG	5136
OY	4795	ggaaaagaaatcacaatgcggtgtlactaagcccgactctctatccactctcttgaagggat	4855
Db	5137	AGAAAAGAAAACACATGTGAATGACTAAACCGATTTTATTCACATTTTCATGATGTATT	5196
OY	4855	ttttctccacatctaaatttcctcgtggcatctccctcctcgtcatcccaatgctgaacctgy	4914
Db	5197	TTTCGCAATTCCTCTTAACCTTCCTTAAGAAATATCAATCTTTAATTAATCACTGTTAATATTA	5256
OY	4915	tacagctcatcataatgatcatgcatg	4939
Db	5257	TACAGTTTCATTTAAAAAATGTAGG	5281

RESULT	15
LOCUS	PENG9195
DEFINITION	P.falciparum major merozoite surface antigen (gp195) gene, complete cds.
ACCESSION	M37213.1
VERSION	GI:160315
KEYWORDS	major merozoite surface antigen. P.falciparum (strain Uganda-Palo Alto) DNA.
SOURCE	Plasmodium falciparum Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE	1 (bases 1 to 5276) Chang,S.P., Kramer,K.J., Yamaga,K.M., Kato,A., Case,S.E. and Siddiqui,W.A. Plasmodium falciparum: Gene structure and hydropathy profile of the major merozoite surface antigen (gp195) of the Uganda-Palo Alto isolate
TITLE	Location/Qualifiers
JOURNAL MEDLINE	1. 5276
FEATURES	/organism="Plasmodium falciparum"
CDS	/db_xref="taxon:5833" 1..5181 /note="Major merozoite surface antigen" /codon_start=1 /protein_id="AA29611.1" /db_xref="gi:160316" /translation="MKIIIFLCSPFLPIINOCVTHESYOELVKKLEALPDAYLTGCVLFHREKMILNEETITGTGASQSGTSGTSRGSSTSGTSAGTSGTSAGTSGTSAGTSGTSGTSRGSNTLPRSKTSSGASPBPADSDDKKSTADLKHYRVNYLFTIKELKYPFLDLTNHMLTLCDNIHGKRYLIDYEETINLLYLNFPDLLRAKLINDVCADYCOIIPENLRIRANELDVLKLVGYRRPLDNIKDNGKMEDYIKKKTTI ANINELLCSKITDONKNADNEGKKRLVQAQYDLSLYNQLSEAHNLISVEKRIDITKKNENKELDLKINEIKNPDPANSNGTPTLIDKMKRIEEHEEKIKELAKTIKENMDSLTEPDLELEYLYREKKNKVDTPRKODPPKSVOIQVPKPNIYPLPLDTIHNSL AADDKSNYSGLDMPTKERINEKIITDNKERKFIINNIRKQIDLEKKINHREONRK KLDEYDRSKKKADYELLEKEFYEMKFNENPKDVDVDTLFAVAITYNVBKORNNKFSSK NNSVYNVOKSLKLDIYLEDYLSRGNISSEKDFPNYYTLTKGLEADIKKLTETKISSERR

Dh 1741 TCATATCTTGAAGATTATTCTTTAAGAAAAGAAATTTCTGAAAAAGATTTAATCATTAAT 1800
Oy 1711 aagaatccataagtaagaatcgaaacagatcgagacgcttggttgagaaatlaagaag 1770
Dh 1801 TATACTTTGAAAACGCGCTCGAACCTGATATAAAAAATTAAACGAGAAATTAAGAAAT 1860
Oy 1771 gatgaagaacagltgttgagagaagaatlacaaagaacgaataaaccagatgagaag 1830
Dh 1861 AGTGAAGAAACAAATTCGTGAAGAAAAATTTTAAAGACTACACATTCAGCAATATG---CT 1917
Oy 1831 atctcgaggttcgcgaatgtttaaagtcgaagtcgagaagtgctccatcgaagaag 1890
Dh 1918 TCCTTAGAAGTATATGATATTTGTAATTAACGATCAAAAAAGTTTATTAATTAATAA 1977
Oy 1891 atgtatgaactaagaagaactcaactcctcgaagaagtgagtgatlaaataatata 1950
Dh 1978 ATAGAAAGACTTAGAAGATAGAAATTAATTTTAAAAAATGACAACTTAAAGATAGTAAT 2037
Oy 1951 catgtccgaatagttataaagcagagaatlaagcagaacacactaactcactc 2010
Dh 2038 CATGTACCAAAATATTTATTAACACCAAAATTAACGAAACATTAATTTATTTATTAATTA 2097
Oy 2011 aagaagaagatagacaactgaagtgctcatgcccgaagtcgagaagcttgatcaagaa 2070
Dh 2098 AAAAAAGAGTATGATTAATTAAGAAATTTATACCAAAAGTAAAGACATGTTAA--GA 2155
Oy 2071 gagaagaagaacactlaaactgaagtcagatlaactccgagctccacagaaga 2130
Dh 2156 AAGAACACAGCTCTTATTCACATTAATTCACAACTTTACTTGCAGACAGCAACCACTG 2215
Oy 2131 gagataaccgagacaggtctacacacagcccgagacacagcggtgtcagctcgcagagc 2190
Dh 2216 AAGAT--GGGGGTCACCTCCACACACACATTAATCCCAATGAGAAACAGAAATACAGAA 2274
Oy 2191 gatagctgtgaagctcgaagcagacagagcagagcagcagctccagtcagtcagtc 2250
Dh 2275 GAAACAGAAAGACAGAAAGAAAGAGAGACACACAAACGATTAACATTAACATTAACA 2334
Oy 2251 gttcagaggtcgaaggtcgaagtgctcactacacacagcctccgtgaatacagaagccgag 2310
Dh 2335 CCAAAAGAAAGTAAAGTGTAAAAATTCAAATGAACATTAAGATAGTATACAAATTCACA 2394
Oy 2311 aatgtcagcaaacctgtacactccttgagaagcctcactgaatcctgaatacactcactc 2370
Dh 2395 GCGTTGACAAAACAGCTTATCTTAAGAAATTAAGATGAATTTTAACTTAATCATATATA 2454
Oy 2371 tgcacaataatataccgtcgtcctcgaagcactgaacagagaagatcctlaaacagtac 2430
Dh 2455 TGTCTAATAATATATTTAGTATCAAACTCTAGTATGACCAAAAAATTTATTAAGATATAT 2514
Oy 2431 aagataaccaagaagagagagatgaactcgtcctcctgtgataccactgagcctgtctc 2490
Dh 2515 AATCTTACTCCAGAGAAAGAAATGAATTAATTAATCATGTGATTCATTAATTTATTAATTT 2574
Oy 2491 aatataccagaacacatccgtctatgtatctatgtcgaatccactcaactcctc 2550
Dh 2575 AATATTTCAAAATTAACATACCTCTATGTATTCATATATATGATATGAACATATGATTTA 2634
Oy 2551 tctcaactgttcatcgaatataatagagaagatggtctgcaacctgtataaactcaaa 2610
Dh 2635 CAACATCTCTCTTTTGAATTAATCAAAAGAAATGATTTATTAATTAATTAATTAATAA 2694
Oy 2611 gacaacgacaagataaagaactcctcgtgagagaagcctaagaagctc 2657
Dh 2695 GAGGAAATACATCAAAAAATTAATTAAGAGCAAAAAAATTAATTAATTAATTAATTAAT 2754
Oy 2658 -----cactctgttaaacctcctc 2681
Dh 2755 ACATCCAGCTCTGGAATTAACAACGTAATTAATCTGCTCAATCCGCAACTACAGTAATTC 2814
Oy 2682 cagctccatcgaacacagctcctcactcaacctcaagacaa----- 2720
Dh 2815 CAAAACCAACATCAAAATCATCTCTACCAATTAACCAAAATGATGATAGCTATATCATCT 2874

Oy 2721 -----gccgaagtgagcgtacacagacacactcactcagaccacttataacta 2775
Dh 2875 GGTCTGCTGTGATGTAAGAAAGTATGATATCCCTTAACAGATTTGTCTATTAATGATACGAT 2934
Oy 2776 ctgaactgtttgagaacactcctcgtcctcgcgaagaatlaagaacat---ctaccaaga 2832
Dh 2935 TTGAAGGATATGTTAGTCTCTTAATCTTGGAATTAACATAAGTACCTAATCCATTA 2994
Oy 2833 cttatgtgcagaagaatcgtccgagaactctcagagaagatctgaaagacgcgagaca 2892
Dh 2995 ACCATTTCTACACAGAGATGGAATAATTTATGGAATATTTTAAAAAATTAATGATACC 3054
Oy 2893 tctataacgagaagctcactcaactcctcgtgaactcctaagccgagatatacaactcct 2952
Dh 3055 TATTTTAATGATATATCAACAATTCGTAATTAATCTAATTTCAAAAGTATTAACAGTTTG 3114
Oy 2953 aacgatgaatctaaacglaagaagctggaagaagacatcaataaagctgaagaagacactg 3012
Dh 3115 ACCGAACACAAAAA---AATGCATTAATGATGAATAATTAATAAATTAAGATACTTTA 3171
Oy 3013 caactgagcttgactcgtgaacaaagtcacaactgaaactggaagagctcctcgacaag 3072
Dh 3172 CAGTATCATTTGATTTATATATTAATAATTAATAATTAATAATTAATTAATTAATTAAG 3231
Oy 3073 aagaagacagctcgcgaagatagaatcagatcgaagaagctcgtcgaagaagcag 3132
Dh 3232 AAAAAAGAACTTGCCGACAGCAAAATGCAAAATTTAAAAAATTACTTTATTAATAAGACAA 3291
Oy 3133 ctgaaagcaaacactcaactcactgaacaaatccgaaacacgtactcgaagaactcagtg 3192
Dh 3292 TTGAATCAAAATGAAATTCATTAATAATCAACCAATTAATGATTAATTAATAAATCTTCTGTT 3351
Oy 3193 tctctcaacaagaagaagacgcgagatcgcgcgagacagagaacactctggaagaacac 3252
Dh 3352 TTTCTTTAAACAAAAAAGAAAGACCTGAAATGACGAAGTGAACACATTAAGAAACAA 3411
Oy 3253 aagatctctcaaacactcaaacgaagcctcgtcgaatataatggcgagctcctcct 3312
Dh 3412 AATAATTTATTAACAACTTTAATGAAGACTTGTAAATATTAATTAATGATATCATCTCA 3471
Oy 3313 ctgaagactcctcgcgagagagacatccagaccgagataactcagccacactcgaagac 3372
Dh 3472 TTTAAAACTTTAAGTAAGATATCAATTCAAACGAAGATTAATTAATTAAGCAATTAAGAAAA 3531
Oy 3373 ttcgaagctcgtcgaactcgaagcgaagctgaagacacactgaactggaagaag 3432
Dh 3532 TTTAGATATTAAGTAAATTAAGATGAAAACTCAATGATTAATTTACATTTAGGAAAGAA 3591
Oy 3433 aagctcagctactcctcgtcagagacgtcgaactcgtatcgcgagctcgaagaagctcact 3492
Dh 3592 AATATATCTTTCTTTATCAAGTGAATTAATCACTGTTAATTAATTAATTAATTAATTAAT 3651
Oy 3493 aagaacaagaactacacccgacaatagcccaagcgaagataatacagaacgtgataacgca 3552
Dh 3652 AAAAAATTAATTAACAGATTAATTCOCAAGTGAATAATTAAGAAATTAAGCAAGCT 3711
Oy 3553 ctggaactctacaagaagctcgtcgtcgaagaagaagatgctgcgaactgtggtc--- 3608
Dh 3712 TTTAAATCTTACGAAATTTCTCCAGAGCAAAAGTATCAACAGTTGATACCTCACCT 3771
Oy 3609 -----tgaatcgtcgcgacaacactggaagcagctc 3641
Dh 3772 CAACGAGATGTAATCTCCATTAATTTGTAAGGGTAAGTGATGTTCCAGGATCCACA 3831
Oy 3642 acctagaag-----ctgcatctactcatgctcgaagccgaagtcacatacaactcacc 3693
Dh 3832 AAGAAGAAACAAATTAACCACTTACAGCTCTTTATTAACGAATTAACAAAGATGA 3891
Oy 3694 acatctcagaacgtcgaagaatgagtcgaatgagtcacatctgtgctcactcctcgagag 3753
Dh 3892 CAATTACAAATTAATGACGAAGATGATTCCTTAGTGTATTCATTTTGGAGAA 3951

[illegible]

Accession	Sequence	Position
Db	GAAGAAATTCCTGAGTGTAGCAACGGAAGAAATTCACATGTGATGTACTTAACCTGATTCCT	5085
Oy	tatccactcttgacggagatlltggctccagctctaattccctggcaatccctctctg	4893
Db	TATCACTTTTCATGTGATTTTCTCAGTCCCTCAACTCTTAACCTTGTAGGAAATATCATTTCTTA	5145
Oy	ctgactccatgctgcgactccgtcagcctctcaatctaatagatcgatg	4939
Db	TTAAATCACTCATGTAAATATATATATACAGTTTCATTTAAAAAATGTAGG	5191

Search completed: October 27, 2001, 13:14:31
Job time: 14107 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2001, 09:20:04 ; Search time 136.83 Seconds
(without alignments)
6807.065 Million cell updates/sec

Title: US-09-269-874-1

Perfect score: 4920

Sequence: 1 atgaagatcatatcttctt.....tattacagttcattaa 4920

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCRTUS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/Backfilest1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	2698	54.8	5181	1 US-08-257-073-10	Sequence 10, Appl
2	151.2	3.1	3095	5 5231168-1	Patent No. 5231168
3	136.4	2.8	4766	5 PCT-US93-07261-10	Sequence 10, Appl
4	135.4	2.8	2223	1 US-08-257-073-4	Sequence 4, Appl
5	132.6	2.7	6152	4 US-08-973-462-1	Sequence 1, Appl
6	132.2	2.7	19124	2 US-08-487-826B-13	Sequence 13, Appl
7	131.8	2.7	5361	2 US-08-973-462-2	Sequence 2, Appl
8	124	2.5	19124	2 US-08-487-826B-13	Sequence 13, Appl
9	118.6	2.4	9636	1 US-08-323-170B-1	Sequence 13, Appl
10	103.6	2.1	5852	1 US-07-867-106-2	Sequence 1, Appl
11	102	2.1	8920	2 US-08-446-855A-1	Sequence 1, Appl
12	102	2.1	8920	4 US-09-150-741-1	Sequence 1, Appl
13	101.4	2.1	3763	1 US-07-792-865D-1	Sequence 1, Appl
14	101	2.0	658	4 US-08-998-416-595	Sequence 595, App
15	97.8	2.0	837	4 US-08-998-416-288	Sequence 288, App
16	96.2	2.0	6243	2 US-09-056-075-1	Sequence 1, Appl
17	95.8	1.9	3926	2 US-08-731-722-1	Sequence 1, Appl
18	95.8	1.9	3926	2 US-08-731-722-2	Sequence 2, Appl
19	95.2	1.9	8457	1 US-07-991-867B-1	Sequence 1, Appl
20	95.2	1.9	8457	2 US-08-544-332-1	Sequence 1, Appl
21	94.4	1.9	1395	1 US-07-991-867B-25	Sequence 25, Appl
22	94.4	1.9	1395	1 US-08-107-755A-25	Sequence 25, Appl
23	94.4	1.9	1395	2 US-08-544-332-25	Sequence 25, Appl
24	94.4	1.9	6768	1 US-08-107-755A-1	Sequence 25, Appl
25	92	1.9	636	4 US-08-998-416-1137	Sequence 1137, Ap
26	90.8	1.8	615	4 US-08-998-416-186	Sequence 186, App
27	89.8	1.8	1511	1 US-07-991-867B-8	Sequence 8, Appl

28	89.8	1.8	1511	1 US-08-107-755A-8	Sequence 8, Appl
29	89.8	1.8	1511	2 US-08-544-332-8	Sequence 8, Appl
30	88.8	1.8	5852	1 US-07-867-106-2	Sequence 2, Appl
31	87	1.8	3933	2 US-08-731-722-3	Sequence 3, Appl
32	86.4	1.8	4673	1 US-07-638-431-1	Sequence 1, Appl
33	86.4	1.8	4673	5 PCT-US92-00018-1	Sequence 1, Appl
34	84.6	1.7	1511	1 US-07-991-867B-8	Sequence 8, Appl
35	84.6	1.7	1511	1 US-08-107-755A-8	Sequence 8, Appl
36	84.6	1.7	1511	2 US-08-544-332-8	Sequence 8, Appl
37	84.6	1.7	51952	3 US-08-947-823-1	Sequence 1, Appl
38	84	1.7	1230	4 US-09-140-466-1	Sequence 1, Appl
39	82	1.7	7218	1 US-08-232-463-14	Sequence 14, Appl
40	80.8	1.6	665	2 US-08-883-795A-36	Sequence 36, Appl
41	80.8	1.6	2447	2 US-09-014-969-14	Sequence 14, Appl
42	80.4	1.6	3701	4 US-08-845-258-10	Sequence 10, Appl
43	80.4	1.6	3701	4 US-08-990-571-10	Sequence 10, Appl
44	80.2	1.6	8700	2 US-08-392-625-16	Sequence 16, Appl
45	80.2	1.6	8700	2 US-08-466-961A-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-08-257-073-10
; Sequence 10, Application US/08257073
; Patent No. 5766597
GENERAL INFORMATION:
APPLICANT: Paoletti, Enzo
APPLICANT: de Talsne, Charles
APPLICANT: Tine, John A.
TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue, 25th Floor
CITY: New York
STATE: New York
COUNTRY: UNITED STATES OF AMERICA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,073
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/075,783
FILING DATE: 11-JUN-1993
APPLICATION NUMBER: US 07/852,305
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,183
FILING DATE: 20-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Frommet, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2570
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 5181 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-257-073-10

[illegible]

Db	3172	CAGTATCATTTGGATTATATATAATAATTAATAATTAAATTGATGATTAATTAAAG	3231
Oy	3064	aagaaacagtttggtaaatlaaabaacaaatlaaanaaacttatttlaaanaaca	3123
Db	3233	AAAAAAGACATTGGCCAAAGCAAAATGCCAAATTAATAAAAACTTATTAATAAGACAA	3291
Oy	3124	ttagaatcaaaattggaatlcacttaataaccaagcatgataccaanaacttcgtc	3183
Db	3292	TTAGAAATCAAAATTCGATTGACTTAATTAACCCCATTAATGTAATTAACAAACTTTCGTG	3351
Oy	3184	ttcttlaacaaaaaagaagcgtgaabaagcgaacgtgaanaacacaltagaanaaca	3243
Db	3352	TTCTTTAACAAAAAAAGAAAGCTGAATATGCGAAGATGAAACACATTAGAAAAACA	3411
Oy	3244	aaaatattattgaaacatataaagacattgtlaaatataaaygggtgaaatctcca	3303
Db	3412	AAAAATATTATTGAAACATTTATAAGCACTTGTTAAATATTATAATGTCGAATCATCTCA	3471
Oy	3304	ttaaacacttlaagttgaagaatcaatccaacgaagatlaattagccagtttgaanaac	3363
Db	3472	TTAAAAACTTTTAAGTGAAGTATCATTTCAACACGAAGATAATTATGCCAATTTAGAAAA	3531
Oy	3364	tttaaglatlaaglaaattagaaggaanaattaaagataatlaattagaanaagaa	3423
Db	3532	TTTAGAGTAAATTAAGTAAATAAGATGGAACCTCAATGATTAATTTACATTAGAAAGAA	3591
Oy	3424	aaatatacacttataccaagttgatacatltaattgtcgtgaatlaaagaatlaa	3463
Db	3592	AAATTTATCTTTCCTTATACAGTGATTAACATCATTTAATTACGAATTTAAAGAACTATA	3651
Oy	3484	aaaaataaaattiraaglaaatttccaaatggaanaataaagaaatgataagaaatgca	3543
Db	3652	AAAAATTAATAATTATACAGGTAATTCCTCCAGGAAATAATAAGCAAGTTATACGAACCT	3711
Oy	3544	ttagaatcttcaaaaaaaattlccccaagaaggaacgaat-----	3582
Db	3712	TTAAATATCTTCGAAAAATTTTTCGCCGAAGCAAAAGTTTACAACAGTTGTAACTCCACT	3771
Oy	3583	-----gttgcacagttgtlaagltgaagttgataccgaacacattga	3624
Db	3772	CAACAGATGTAACTCCATCTCCCATTAATCTGTAAAGGGTAAGTGAAGTTCAGATCCACA	3831
Oy	3625	caagtgcaaccgaagaacccaagcataactcaltgtagggcgagttccacaacataca	3664
Db	3832	AAAGAGAAACACAAATTAACCAACTTCAAGCTCTTTATTAACGAATTTACAACAGTAATA	3891
Oy	3685	acatcacaaaatgctcgatgtagaagatgaacgtatcatagatccatatttggagaa	3744
Db	3892	CAATTACAAAATTAATGACGAAGAAGATGATTCCTTAAGTTGATTTGCCATTTTGGAGAA	3951
Oy	3745	tcgagaagaagattatgataatttaggacaagtatcaacagagaagcgatlaactcctcc	3804
Db	3952	TCCCAAGATATATGCGAATTTTATGATCAAGTAAGTAACGTGGAGAGCATATA---TCGTGC	4008
Oy	3805	gtaattgataacacacttctaataattgnaaaltgaaatgaggttttatatttlaaacct	3864
Db	4009	ACAATGGATATATCCCTCCTAGGATTTGAAAATGAATGATGATTAATTTAAACCT	4068
Oy	3865	ttagcaggttcttataagaagtctlaaaaaaacattagaanaataagttatgacattaat	3924
Db	4069	TTTACTGGAGATATATGAAACCTTTAAAAAACAATTTGAAAAAACAATTTTACATTAAAT	4128
Oy	3925	gttaagtcttaagatattttaaattcaagatltlaaactlaaaccgtlgaanaattccaanaatgct	3984
Db	4129	TTAAATTTGAACGATATCTTAAATTTACGCTCTTAAGAAACGAAATATTTCTTGAGATGA	4188
Oy	3985	ttagaatcgaatttaattcatataaagaatttaacataagaatltatgctgtcaagaat	4044
Db	4189	TTGAATATCTGATTTTAATGCATTTTAAACATATATTCCTCAAAAGATACATTTATGAAGAT	4248
Oy	4045	ccatataaattcttlaalaagaanaaagaataaattcttlaagcagttataatataat	4104

Db 4249 TCATTTAAATTTGTAATTCGACCAAAAAAGACACTTTTAAAGTTACAAATATATA 4308
QY 4105 aagattcaatagatocgagatataaatitctgaagaatgctcttgatataataata 4164
Db 4309 AAGGATTCAGTAAATAATGATTAATTTGACAGAGAGATTAAGTTTATGAAAAG 4368
QY 4165 tlatcgaataataataatagattagattcaataataataataataataataataata 4214
Db 4369 GTTTTGGCCGAATATTAAGATGATTTGATTAATTAAGTTTAAAGTTTAAAGAGAAAG 4428
QY 4215 -----cgacaacaa 4224
Db 4429 GAGAGTTCCCATCATCCACCAACACACCTCCGTCACACGACAAAACAGACGACAA 4488
QY 4225 gctgaataagagaataatcctccctctttaaataatagagacattataaacaagtc 4284
Db 4489 AAGAGGAAGTAAAGTTCCTTCCATTTTAAACAACTTGAAGCCTTATACAAATACTTA 4518
QY 4285 aatgataaaattgattatctgaattcattagagaagcaaaagttcaaatatatacat 4344
Db 4549 GTTAATTAATTTGACGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 4608
QY 4345 gagaatcaaatgagagatataataataaagaacttaattcaataaacaatcaaac 4404
Db 4609 GAAAGAGATGAACACATCTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 4668
QY 4405 aatttgacagatttaaaaaataaataaacttctgagagagagagagagagagagag 4464
Db 4669 AATATAGATCTTTTAAACCAATTAAGCATTGGAAGCAATTAATAATTAATTAATTAAT 4728
QY 4465 tataacataactatctgacaaagttccttgatagagagagagagagagagagagag 4524
Db 4729 GATACGAAAAAATATGCTTGGCAATTAATTAATTAATTAATTAATTAATTAATTTT 4785
QY 4525 gctaaaaacgtttatcttaattcttgatagagagagagagagagagagagagagag 4584
Db 4786 CTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTCA 4845
QY 4585 caac 4644
Db 4846 CAACACCAATTCGTAATAAATCAATTCGTAATAAATTCGTAATAAATTCGTAATAAAT 4905
QY 4645 gaaag 4704
Db 4906 GAAAGAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 4965
QY 4705 aatccaaatctctcttgatagagagagagagagagagagagagagagagagagag 4764
Db 4966 AATCCAAATCTCTCTGTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTT 5025
QY 4765 gaagaag 4824
Db 5026 GAAAGAGATTCAGTAAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5085
QY 4825 tatccacattctgag 4884
Db 5086 TATCCACTTTTCAGTAAATTTCTGCAAGTTCCCTAATCTTTAGAGAGAGAGAGAGAG 5145
QY 4885 ttaataactatgataataataataataataataataataataataataataataata 4920
Db 5146 TTAATCTCATGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAA 5181

RESULT 2

5231168-1
; Patent No. 5231168
; APPLICANT: DZIGIEL, MORTEN; BORRE, MARTIN; JEPSEN, SOREN;
; VUUST, JENS; RIENECK, KLAUS; WIND, ANNETTE; JAKOBSEN, PALLE H.
; TITLE OF INVENTION: MALARIA ANTIGEN
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/409,658
; FILING DATE: 18-SEP-1989

; SEQ ID NO:1:
; LENGTH: 3095
5231168-1
Query Match 3.1%; Score 151.2; DB 6; Length 3095;
Best Local Similarity 42.9%; Pred. No. 4,5e-16;
Matches 1310; Conservative 0; Mismatches 1688; Indels 56; Gaps 9;
QY 512 atgaagaataatgaatattatataataataactttttttgtatttttaagcaaa 571
Db 47 aagaagtagcttgaagaagggccatcagaagaattgacctgaaacaataatgaag 106
QY 572 aattaaatgctatgctgaatgatttgcataacttccatcttaaatctgag 631
Db 107 aatcaggtgaagaataatattgtgataatgaagaagtggttttgagaagccatcag 166
QY 632 caaatgaattagacgtacttaaaaacttgctcgatataagaacaattagacata 691
Db 167 aaaaatttcatctgaaagtagtaactcgtgaatttaattgaataatgaattgttgaa 226
QY 692 ttaagaataatgagagaaaatggaagatctacattaaaaataataaacaatgaaga 751
Db 227 acaaaagtgaactgaacctgctgaacatgaagaagttgattctgagaagaagcaacct 286
QY 752 atataatgaattaatgaagaagaatgaagaacaattgataaataatgaatgcaacta 811
Db 287 aaccagctgaagaataatgaagaagtagtataagaagaagctctcaggaagaattgtact 346
QY 812 aagaagaagaaaaaaaatataccaagctcaataatgacttccattataaataac 871
Db 347 aacaaaatgtagaagatcagtgtagaagtgatagttagtataatgaagaagtgatttg 406
QY 872 aattagaagaagcacataatttaataagcgttttagaagaacgatgacatttaaaa 931
Db 407 aagaacctaactatgaagaattggaacctgatacaaatgacctgaattgaatgaagaat 466
QY 932 aaaaatgaacatlaaggaatctactgataagaataatgaataatcccccccg 991
Db 467 aattagtgatcagaagaagaagtgatctgaaccagctgacatgtagaattgtatcag 526
QY 992 ccaattctggaataacacccaactctccttgataagaacaaaaaatcgaagaacag 1051
Db 527 aaaaagtgatctgaaacagcagtagaagaattgattctctgaaaaaagtagacacg 586
QY 1052 aaaaagaataaagaagaattgccaacacttaatttaattatgtatgtatttactg 1111
Db 587 aaccagctgacatgtagaagtgatctgaaacaagtagaataacgaaccatccgaaga 646
QY 1112 atccactgattagaataactatttaagaagaaaaaataaataatgatatagtgcaa 1171
Db 647 aagaatgacaggtctcctcaaaacattgaagaagaattgaagaagtgagatgccaacta 706
QY 1172 aggttgaacaaagaatcaactgaacccaatgaatatccaatgaggttactatcct 1231
Db 707 aatttgagaccttcaataataattgaacctaatlltggctcacaacccaatccaaag 766
QY 1232 tgcataatacgatataaacaatgctttaaataatgaacttaattcttgggtatttaata 1291
Db 767 aaccagttgacacatcatlgtlcaaaaattgaaaagttcctctgagaagaataaacaat 826
QY 1292 atccatttgattatacaaaagaaccaaagtaaaaactatatactgataatgaagaanaa 1351
Db 827 caagtggttgcctcgaagtagaagaagaagaagaatgatactcgaagtggttgaagaanaa 866
QY 1352 aattcataatgaatgaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1411
Db 887 aaaaatcacaagaatcagtggaagaagaattccagtaaatgaggttgatgtgaagaagttc 946
QY 1412 aaaaatcttaagaagaagatcgaagttttaaataatgatatatacaaaagaatgaagaat 1471
Db 947 acactgaacaaattagatttagatcataaaacagtgatccagaagaatgtagaagttgaag 1006

QY	1472	tacttaatgaaattatgatagcgaatctcaataataataataatgattttaaacttaatttcggaa	1531
Db	1007	aaattctctcagaactcactcagaataatgaaatggtctcattccagaataattgtgaattggg	1066
QY	1532	aaatgattgggtlaaagatcatctcatlaaag---ttgagaacttaccacccaataactt	1588
Db	1067	aagttcttccttgagacccaatacaaaataacgaaatttcaaagaataatgaaatgtgataaa	1126
QY	1589	ttgcattcctatgaaataattctcaaacataacttggaaagttacaagaagcttcaatata	1648
Db	1127	gtgcacatacttcgcattgaaatagtatgaaagtagaaagaaataactctccgaagatgataa	1186
QY	1649	tggagaattacttcttaaggaatatagtatgttgaanaagaatlaaatafatataaanaat	1708
Db	1187	atgnaaagaattgaaacatgaaatagtagaagttgaagaaattctaccgaagataaanaatg	1246
QY	1709	taataagcaaaatcagaa-----atgagattgaaacattagtgtgaanaattat	1786
Db	1247	aaaaagctcaacatgaaatagtagaggttgaagaaattctaccgaagatgataaanaatg	1306
QY	1757	aaaaagttgagaagacgcct-----tttggaaaaaaaattctcaagcagaataaac	1810
Db	1307	aaaaagttgacactgaaatcgtatgtagaaggttgaagaaattctaccgaagataaanaatgaa	1366
QY	1811	cagatgaaaaaattttagaagtaactgcacattgtlaaagttacaagtctcaanaagtttat	1870
Db	1367	aaggtcaacatgaaatagtagaggttgaagaaattctaccgaagataaanaatgnaaag	1426
QY	1871	taatgacaaaatttgcacgaatttaaaaagactcaattgattttaaaaaatgtagaattaa	1930
Db	1427	ttgacacatgaaatagtagaaggttgaagaaattctccagaagatataaattgaaaaggtc	1486
QY	1931	aacataatacactgttcccaactcttacaacaagaanaatagaaagaaacttatatt	1990
Db	1487	aacttgaataatagtagaggttgaagaaattctcacgaagataaanaatgaaaagttcac	1546
QY	1991	taattgttttgaanaaagaataatgataaataaagttgttcactgactaggtatgatat	2050
Db	1547	atgnaatagtagaaggttgaagaaattctcacgaagataaanaatgaaaaggtcacaatg	1606
QY	2051	tgtataatctgagaaaaaaaatacaataaaaacgaaggttcaatcgtatataatccgaacat	2110
Db	1607	aaatagtagaggttgaagaaattctcacgaagaagatataaataatgaaaaggtcacaatg	1666
QY	2111	caactcgaagagagaataatacagggacagacactcaacaaacttggacaacaagcagatctg	2170
Db	1667	aaatagtagaggttgaagaaattctctccgaagataaanaatgaaaaggttcaacatgaa	1726
QY	2171	ctttagaaggaagattcagtaacaagcacaagcacaagaanaaacaagcacacaaccag	2230
Db	1727	tagtagaggttgaagaaattctcacgaagataaanaatgaaaaggttcaacatgaaatag	1786
QY	2231	taccagatcacgtacacgaagcagaagcacaagttcccaacacccacgcacacgtataata	2280
Db	1787	tagaggttgaagaaattcttcacgaagaattgttgaattgaagaaatgaccatcacaaca	1846
QY	2291	ataaactctgaanaatgttcccaattgatactcttgaanaatattatgaatttttaata	2350
Db	1847	atacaatgaaataatttgaacataataaacacgaagaanaaaaaagaatgttaatttagtctg	1906
QY	2351	cttcataataatgtccaaataatattttgtttccacactcaactatgacgaanaagatat	2410
Db	1907	aagaaaaagcaattccacagaacccgttgtaacctcatltaaaatgaaaaatgaaaacgtta	1966
QY	2411	taaaacataataaattatcagaaggaaggaagaaagcacaattaaagttcatggtatccatag	2470
Db	1967	ctcccaaacacatctgaaaggtgaaatccactaacaacacgaatatagttcaaatataaatagtac	2026
QY	2471	acttattgttttaatacaanaataacataactccgttaatgattctctgtttgatagcttaa	2530
Db	2027	aagaaaaataaccaataaanaaagaaaccccgtagtagatggtgtccaaaacgttgaagac	2086
QY	2531	acaaatgatttatcacaaactatttatgaaattatgatbaaagaagaatgtgttgtaatttat	2590

[illegible]

```

RESULT      3
PCT-US93-07261-10
; Sequence 10, Application PC/RUS9307261
; GENERAL INFORMATION:
; TITLE OF INVENTION:  P1EMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF

```

```

: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: John H. C. Blasdale
: STREET: One Giralda Farms
: CITY: Madison
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07940-1000
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: Apple Macintosh
: OPERATING SYSTEM: Macintosh 6.0.5
: SOFTWARE: Microsoft Word 5.1a
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/07261
: FILING DATE: 19930805
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 07/927,531
: FILING DATE: 07-AUG-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Blasdale, John H. C.
: REGISTRATION NUMBER: 31,895
: REFERENCE/DOCKET NUMBER: DX0288K
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201-822-7398
: TELEFAX: 201-822-7039
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4766 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: ORIGINAL SOURCE:
: ORGANISM: Plasmodium falciparum
: STRAIN: Malayan Camp
: IMMEDIATE SOURCE:
: CLONE: p2b1;p12-1
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 3..4766
: PCT-US93-07261-10

```

Query Match 2.8%; Score 136.4; DB 5; Length 4766;
 Best Local Similarity 42.1%; Pred. No. 1e-13;
 Matches 1577; Conservative 0; Mismatches 2111; Indels 54; Gaps 12;

```

Oy 515 aagaaataagaaataatataataaataaactttatttgaattatataagaagcaaat 574
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 244 AGGAATAAGAAAGAAATAAATAAAGATTAATGATGAAGCATTAATAAGATTAAG 303
Oy 575 taatgatgtatgtctaatgtattgtcacaatacccttcaatcttaaatctgtgcaa 634
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 304 GAAATGATTAAGATGATTA-AAAATAGTCCCAAAAAACACAGAAATCCGTAGAAAAGAT 362
Oy 635 atgaattagacgtacttaaaaacttgcgttcgatatagaanaaccattagacaatatla 694
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 363 TTGCAAGAATAATGAATTTGAAGAAGAGAAATTTATAAAACACATTTAAAGATTGAG 422
Oy 695 aagataatgtagaanaaaltggaagattacatataaanaaaataaanaaacatagaanaata 754
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 423 GAGAGAAAGAAAGAAAGAAATGATTAAGAAAGTTTAAGAGAGATAATTAATTAAGA 482
Oy 755 taatgaattaatgtagaanaagtaagaanaacaattgataaanaagaatgcaactaaag 814
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 483 GAAATTTGAACATTTAGAAAGTTTAATGCTCAATTTGAAGAGTCTTAATGAATGAATTA 542
Oy 815 aagaagaaaaaanaaataatataaccagctcaatgataagcttcttatttacaataaacaat 874
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 543 GAAAGAGAGAGCTTTCAAGAAAGACCTATGATGTTTAATGCAACG--TGAATGAAGAAATG 600
Oy 875 tagaagaagacataaatttaataagcgttttagaanaaacgtattgcaactttaaaaaaa 934

```

```

Db 601 AACTTGATGAATGATTAATAAATAATGATGATGAACAGACAGAAAGAAATGTACGAAG 660
Oy 935 atgaanaacattagaagattacttgataagaataaagaattaaatccccccggcca 994
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 ATGAGAAATTAAGATTAAGATGATGTTATGACAGAAATTTGGAACCAATTTTATG 720
Oy 995 attctgnaatacaccnaatctcccttgataagaacaaataatcgagaacagaa- 1053
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 GTATGAGAGAAATGCACTTTGGCAATTTAGATGAATATGAGAAAGATATGAAAAAAC 780
Oy 1054 -aaagaaataaagaanaaltgccaanaacttaataatattgtagtattactg 1111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 GTTATTATTAAAGAAAGATGAGAGCGTGATTTAAAGATGTTGAAGAAATTTAGAAC 840
Oy 1112 atccacttgattagatactatttaagagaanaaataaataatattgataagtgcaa 1171
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 AAACGTGTTATGTTTATGAGAAATAATTTCTTCAACACAGATATTTGTTAAAGAAAG 900
Oy 1172 aggttgaacaaaggaatcaactgaaccacatgaatatcccaatggagttactatccct 1231
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 901 GAATTAAGACAAAAAACTTAAGAGACACAAAGAAAAAATTAATTTGCTGCTGAC 960
Oy 1232 tgcataatacgaatatacaaatgctttaaagttaacttaattcttgggtgatttaata 1291
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 961 AACACAGCATGAGAAAAAATTAATTAAGATTTCTGATGATAAGTTGTGTTCCCG 1020
Oy 1292 atccatttgattacaaaagaacaaagtaaaaacatatatcttgataatagaagaaaaa 1351
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1021 TGAATTAATAATTAATCATCTTTCTGATTAATTTAGAGCTCCAGATTAATAAGAACGA 1080
Oy 1352 aattcataatgaaatlaagaaanaaataaataagaanaaanaaattgaaatcgata 1411
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1081 TGTTTATAGACTAAGATTAATTTCTTCTTATTTGACCAAGAAAAAGATATGACCTTGCTG 1140
Oy 1412 aaaaatcttagaagaacagatctgaagctttaaagataaacaagaatatgaaanaat 1471
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1141 TGTGTGTT-----GATAGTATGATTTCAAGAGTAAATGAAAAAATTTGAATGCA 1190
Oy 1472 tacttaatgaattatgtatgaacaaatctcaataatataatagatttaactaatctgaaa 1531
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1191 ACTTTCAATCCGTTTA--AAAGAGAGCGTAATTAACCTTAAGAGAGAAAAATTCACGAAT 1248
Oy 1532 aaatgaatgaaatataatataatataatataatataatataatataatataatataat 1591
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1249 TACACAGATTTAAATAAATTTAATAAATAATCAAAAGTTTGAAGAGAAAGAAAGAG 1307
Oy 1592 catcctatgaaaaatcttaaacataatcttgaagaagttaacaaagctcttaaatatag 1651
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1308 -----GAAATTCAGATGAGTGAACCTTTGAATACACCAAGAAATGATGTTAAGGC 1359
Oy 1652 aagatatcttcttaagaatatagtgcttgaaagaagaatataatataatataatataat 1711
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1360 CTTCAGATTTTAATGATTAAGAGAAATAATTAATCAACAGCTCATCAATTTAATCAAC 1419
Oy 1712 taagcaaaaatagaanaatgagatgaaacattagttaaatatttaaaaagaatgaaagac 1771
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1420 CACCAAAAGATTTAAAGAAATTCAGAAAGAAAGTATGCTCTTAAGATTTATCAATTTGAAC 1479
Oy 1772 agcttllgaaaaaanaatcttaagaacgaaataaaccagatgaaaaaatttagaag 1831
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1480 ATGAAACACCAACCAAAATTAACCTGAATACGAAAGAAAGTATGCTCTAGGAATATCAAC 1539
Oy 1832 tattgacattgttaagaagtaagttcaaaaagtttattatagaaacaaattgacgaat 1891
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1540 TTGATATGTAACCCCAACTTAATTAATTCGTAATATGAAGAGGTCTATGATGAGAAAT 1599
Oy 1892 taanaaagctcaattgattttaaaaatgtagaatttaaacataatataacatgttccca 1951
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1600 ATCACTTGATTAATGAGTACGAGATGAATTTACCTTAATATGAAAAGGTGATGTCTTA 1659
Oy 1952 attcttacaagaagaanaatgaagaagaaacttatatttaattgttggtaaaaaaaga 2011
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


Db 1660 GAGAAATTCACACTTGCATTAAGAGTCCATCTACATTAAGGAATATGATCAAAACGAAT 1719
QY 2012 ttgataaatlaaagtggttcacatgacctaaagtgagaatcattgataaattgaagaaaaaa 2071
Db 1720 TAGCAAAAGGTAAGATTAATACCAACCAACGAAATCCGTTGATGATTAATGATCAAA 1779
QY 2072 acataaaaacagaagtcacatcgataatcggaaacatcaaacggaaggagaatcacg 2131
Db 1780 CAGAAATTACCAAAAGTAAGATTAATACCAACCAACGATGAATCCGTTGATGATTAATG 1839
QY 2132 gacaagcaactcaaaaacctggaacaaacgagatcgtcttagaaggagatcaglac 2191
Db 1840 ATCAATCACAAAATTAAGCAAAAGTAAAGATTAATCACTAACCAACCAATGATCCGTTGATG 1899
QY 2192 aagcaacaagcacaagacacaaaacaaagccacaccagtlaccagtaaccagtaaccgaag 2251
Db 1900 AATATGATCAAAACGAAATTAAGCAAAAGTAAAGGATTAATTAACCAACGAAATTC 1959
QY 2252 caaaagcacaagtcaccaacacacacacagcagcaglaataataaactggaatcgttcca 2311
Db 1960 TGGAAAGAAATACATGAATGAAATGTTAGCAAAAGTAAAGGATTAATTAACCAACCAATG 2019
QY 2312 aattagattacatcgtgaaaaaattatagaaatttlaaatactcatalatagtcacaat 2371
Db 2020 AATCCGTTGATGAATATGATCAATCAATGAAATTAAGCAAAAGTAAAGATTAATTAACCAAA 2079
QY 2372 atatttggcttccacatcaactatgaacgaagaataltlaaacaatatlaaactacaa 2431
Db 2080 CACACGAATCCCTTGATGAATATGATCAAAACGAATTAACCAAGGTAAGGAATTAATCA 2139
QY 2432 aggaggaagaagaagaacttaagatcgtgacatcagatcagatcgttgaatatacaaa 2491
Db 2140 ATTAAGACACGCAAAATCTAGAAAGATTAATGAATCAATGTTAGCAAAAGTAAAGGAAG 2199
QY 2492 ataacataccgttaatgtaattcctaigtgttagagcttlaaacaatagtltaacacaat 2551
Db 2200 TACTATTAATTAACGACCGCAAAATCTGAAGAAATTAATGAATCAATGTTAGCAAAAGTAA 2259
QY 2552 ttatgaaattatgaagaaagaatggttgaattatataaacttaaggatlaagaa 2611
Db 2260 AGGAATTAATTAATTAAGGCAATGAATTCGGAAGATTAATTAAGCAATGTTAGCA 2319
QY 2612 aaattaaaaattatagaagaagcg---aaaaagtlaccacatcgttgaataactctt 2668
Db 2320 AAGGTAAAGGATTAATTAAGGCAATGAATTCGGAAGATTAATTAAGCAATGTTAGCA 2379
QY 2668 caagtltcaatgcaacccatcatalcaacaacctgaagataaaccgaagtgcga 2728
Db 2380 TAGCAAAAGGTAAGGATTAATTAAGGCAATGAATTCGGAAGATTAATTAAGCA 2439
QY 2729 atgatgatcatcacatcctacaaatttgaataatagttlaaataattattgaacaatat 2788
Db 2440 CTGATTTAGCAAAAGGTAAGGATTAATTAAGGCAATGAATTCGGAAGATTAATTAAGCA 2499
QY 2789 tgaagtttgaagaaacaaataatataccaaagatlaaagtlcaaaaagtgcga 2848
Db 2500 ATGAAGAACTGATTTAGCAAAAGGTAAGGATTAATTAAGGCAATGAATTCGGAAG 2559
QY 2849 accttataagaagaatataaagaatagtlgatacatcttataatgaatcctttacaatt 2908
Db 2560 AATATTAATTAAGGATTAATTAAGGCAATGAATTCGGAAGATTAATTAAGGCAATTCGGA 2619
QY 2909 ttgttaaatctaaagctgtgatat-----taattcatgtgaatgtatgaataaaa 2959
Db 2620 TAGAAGAAATTAATTAAGGATTTAGCAAAAGTAAAGGATTAATTAAGGCAATTCGGA 2679
QY 2960 ggaagaatlaagaagaataltlaaataaataaacttaccagttatcatcattgatt 3019
Db 2680 AAAATCTAAGAAATTAATTAAGGATTTAGCAAAAGGTAAGGATTAATTAAGGCAATTCGGA 2739
QY 3020 tatataataataaataaataaagattatctgataaataaagaacagttgta 3079
Db 2740 CACGCGA-AAATCTAGAAAGATTAATTAAGGATTTAGCAAAAGTAAAGGATTAATTAAGCA 2798
QY 3080 aatataaatgcaaatlaaaaaacttacttataaagaacaaatlaagaatcgaattga 3139
Db 2799 AATTAAGGACGCGCAAAATCTAGAAAGATTAAGGATTAATTAAGGATTAATTAAGGATTA 2856
QY 3140 attcacttaatacccccaagcgtatlaaaaaacttctcgttcttlaacaaaaaa 3199
Db 2857 AATTACAAATTAAGGATTCGATGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTA 2916
QY 3200 aagaagctgaatagcagaacactgaacacatlaagaacacacaaaaataltatgaac 3259
Db 2917 TAGGAATTAAGGATTCGATGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTA 2976
QY 3260 attataaagacttgaattatataatgtygaatcattccattaaacttaag 3319
Db 2977 GAATTAAGGATTCGATGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGG 3036
QY 3320 aagaatcaattcaaacagaagaatlaatcgcagttlaagaacacttlaaagtltaagta 3379
Db 3037 ATTAAGGATTCGATGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTA 3096
QY 3380 aattagaaggaaaattaaagataattlaaattlagaagaagaataatlaactat 3439
Db 3097 AAGGATTCGATGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 3456
QY 3440 caagtggattacatcattatgctgaattgaagaaglaaataaataaataatata 3499
Db 3157 GATCTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTA 3210
QY 3500 caggtatctcccaagtgaaaataaataaagatgttacaatgcatcagatcctacaata 3559
Db 3211 AAGGATTCGATGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGG 3270
QY 3560 aatttcccaagaaagacagatgtgcacaggtglaagtgaaatgtaaccgacat 3619
Db 3271 GATCT-----GAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGG 3324
QY 3620 tagaacaagaatcacaagaagaacagcatcagatcagtaggaagcaggtcacaaca 3679
Db 3325 AAGGATTCGATGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGG 3384
QY 3680 taacaacatcacaaaaatgcatgtagaagtagatgcgtlaacatagatcatalattg 3739
Db 3385 GATCTGATGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTA 3444
QY 3740 gagaatccgaagaatltatgatttaggacagtagtaacaggaagcagtaactc 3799
Db 3445 CTGATGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTCG 3504
QY 3800 ctccgtaattgtaacatcattcctaaatltgaagaatgaatltgaaggttlaattaa 3859
Db 3505 AAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGG 3564
QY 3860 aaccttgaagctgttctlaagaagttlaaagaacaaactlaagaataaactgtatgacat 3919
Db 3565 AAGATTTATCTAATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGG 3624
QY 3920 ttaagttaatgltlaaggaatltlaaactcacgattlaaagcgtlaaacttcaaaa 3979
Db 3625 AAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGG 3684
QY 3980 atgtttagaatcagtttaactcattaaagaattlaaagatlaaagtaagttaagtgtca 4039
Db 3685 AAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGG 3744
QY 4040 aagatcatalaataatttctlaaagaagaagaagataaattcctaagcagttacat 4099
Db 3745 AACCAAAATCTGTTTAAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGG 3804
QY 4100 atacttaagattcaatagatagatataaatttgcacaaatgagtggttcttgatata 4159
Db 3805 ATACACCTAATGAAGAACAAACAAATACCTGTTTAAAGGATTAAGGATTAAGGATTAAGGATTA 3864

Db 1472 TGAACAGTATTAAATGATGATGAATAATGATTCTTTGAAAGAAATTAATCGAAA 1531
OY 1577 accataactcttgcacccctatgaaatctcaacataactctgaaagttacaagaag 1636
Db 1532 ATAAATAAACACCTAATGTTGACCAACCAATCAAAAAAAGAAATGAAGAA 1591
OY 1637 ccttaataatactgaaagatcttcttaaggaataagtagttgaaagaataaataat 1696
Db 1592 CTGTTCTCGTATGATGAATAATTTTGAATATCATCTTAAATTTTAAAGAAAT 1651
OY 1697 attataaataatttaagcaaaatagaaatgaaatgaaatgaaatgaaatgaaatgaaat 1756
Db 1652 ATTATTAATGATGAATAATGATGATGAATAATGATGAATAATGATGAATAATGATGA 1711
OY 1757 aaaaagatgaa-----gaacgcttcttgaaaaaaattactaagaagcga 1804
Db 1712 AAAAATTTGACACTTAAATAATGAAATGTTAGTGAACCATTAATTTGATTTAG 1771
OY 1805 ataaacagatgaaataatttagaagatctgacatctgaaagtaacaagttcaaaag 1864
Db 1772 GAAATTAAGATTAATAACCATTTGAAGCTATTAACACAGATTTTCAATTTAATTAACAG 1831
OY 1865 tttaataatgaaacaatactgacgaatttaaaaaagactcaattgattttaaataatgtag 1924
Db 1832 AATTACAAGCTATTATATGAACTTATGAACTTATGAACTTATGAACTTATGAACTTATGAA 1891
OY 1925 a---attaaacataatacagatcttcccaattcttacaacaagaataatgaaagaaac 1981
Db 1892 AATTTTCAACAAATATTTTGAATAATGATGTTCTTATCATCAAGAACGGAAGAA 1951
OY 1982 ctattattattcttgcttgaaaaaagaattgataaataagtgcttcagcctaag 2041
Db 1932 TCGAATAACAACTTAAAGCAATTCACCAACCAATAGAACCTGAAATGATGCCCTTCGCAC 2011
OY 2042 tagaactcttgaataatgaaagaaaaaacaataaacaagaagtcacatcgatatt 2101
Db 2012 CAAAAAATTAAGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2071
OY 2102 cggaaacactcaacggagagaaataacagagacgaactcaaaactctgacaaag 2161
Db 2072 AAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2131
OY 2162 caggatcttgccttagaagagatctgataagcacaagcacaagaacaaaaaagaagc 2221
Db 2132 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2191
OY 2222 aaccacagttaccagttaccagttaccagttaccagttaccagttaccagttaccagtt 2250
Db 2192 CAGAAATTTGACAACTGAGAAATCAAA 2220

RESULT 5
US-08-973-462-1
: Sequence 1, Application US/08973462B
: Patent No. 6191270
: GENERAL INFORMATION:
: APPLICANT: DROULHE, PIERRE
: APPLICANT: DAUBERSIES, PIERRE
: TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
: FILE REFERENCE: 0660-0125-0 PCT
: CURRENT APPLICATION NUMBER: US/08/973,462B
: EARLIER FILING DATE: 1998-02-06
: EARLIER APPLICATION NUMBER: PCT/FR96/00894
: EARLIER FILING DATE: 1996-06-12
: EARLIER APPLICATION NUMBER: FR 95/07007
: EARLIER FILING DATE: 1995-06-13
: NUMBER OF SEQ ID NOS: 29
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 6152
: TYPE: DNA
: ORGANISM: P. falciparum

US-08-973-462-1
Query Match 2.7%; Score 132.6; DB 4; Length 6152;
Best Local Similarity 44.7%; Pred. No. 4,2e-13;
Matches 1320; Conservative 0; Mismatches 1569; Indels 66; Gaps 18;
OY 582 tctatgcttaatgattatctgcaaatcttcaatcttcaatcttcaatcttcaatcttcaatctt 641
Db 2686 tcttctgcaaaacgcttgcacaataattatctcagacacattttaaagattatcttgaag 2745
OY 642 agacgtacttaaaaaactctgcttgcgatatagaaacccatitagacaatatlaagata 701
Db 2746 tctcgaactcgggaataaagaagacagtatattaaatgatatagaagaagtaagaaga 2805
OY 702 tctaggaataatcgggaattacatlaaaaaataaaaaaacocatagaataatataatga 761
Db 2806 tctagtcaccacaatcttgaagaacgtagaagaactcagaactggaagtgtaactctt 2865
OY 762 attaatgagaagaatgaagaacaattgataaaataagaatgcaactgaagaaga 821
Db 2866 tagtaacatatagagagagatacaaaaaataactatctactatgatacattagaagaaga 2925
OY 822 aaaaaaaattatatacgaagctcaatgatcttcttattcaatacaaatlaagaaga 881
Db 2926 attagagaagcctccagaagaatgatatagtgccgctttagaataatacccaagtgaga 2985
OY 882 agcacataatttaataagcgtttagaagaacgtagatgacactttaaanaaataatga 941
Db 2986 ggaagaagaagaatgatatagatctgataatgaagaagaagtgctaccactt 3045
OY 942 ctttaagagattacttgatagaataatgaataataatacccccgcggcaattctg 1001
Db 3046 aatgaagaactctggaacagcgagaagaagaagcgcaatacattacggaatatctga 3105
OY 1002 aatatcaccaactctctcttctgatagaacaaataacggaacacagaagaagaaga 1061
Db 3106 aattttagaagaagaatgacgtagaagaatgaataatgctgcaggaattttagagaatt 3165
OY 1062 aaaaagaattcccaaaactataatata-----ttgagtttatttactgattcc 1115
Db 3166 aaaaagaactctgatttataactctgattagataagtagagaacagtgaaattagcgg 3225
OY 1116 acttgattagatactattttagagaagaataataataattgattagtagagaagtt 1175
Db 3226 agaaagtttagaagaacaagaatgataagcaattttagagaatatctgataatctg 3285
OY 1176 tgaacaagaagaactcaact---gaaccatgaatatccaaatggagttactatctt 1232
Db 3286 aaaaagaatacagaagaatttatttaacaggtatgcttgcgaatagataaaccagttatg 3345
OY 1233 gtcataatacgaatatcaaatgctttaaataatgaacttaattcttctggtatttaata 1292
Db 3346 aatccaatcagaagaagaagttgatttgaatga---aatggttaggttcgaatttaga 3402
OY 1293 tccattgatttacaagaagaacgaagtaaaaaacatatactctgataagaagaagaa 1352
Db 3403 taataagaataatagaaagaagttatttaataaattagaanaattacagaactcga 3462
OY 1353 attcataatgaaatlaagaagaataaataaataaataaataaataaataaataaataaata 1412
Db 3463 aggtgttcaagaactcgttaactcgaacatctgaacacaaatgatatatggtatgattgt 3522
OY 1413 aaaaacttcaagaagacagatcttaagttctttaaattgatacaaaaagaatagtaaaat 1472
Db 3523 tccgtgatgaaagatcaactttagaataatttaaatgagcgaggggttgaagaagaat 3582
OY 1473 acttaatgaattttagatagcaaatcaataataatataatagatttaacttaacttggaaa 1532
Db 3583 gtttttaatttgaagaat---gtatttaagaatgaaatgatatgtaattactcgttagaga 3639
OY 1533 aatgaatggaataagaatatcatataaagttgaagaactcaacacacataacttctgc 1592

Dh 3640 aatlaagatgaacccggttccaaaagatgagaaaaaagaacactctagtattatgtgaaga 3699
Qy 1593 atccctatgaaatcttcaacataatcttgaaaggttaacaaaagctcttaatatatgga 1652
Db 3700 aatggaagaataatctgtgagctgattagaggaagaagaagatttaacagacaagat 3759
Qy 1653 agattcttctaagaat---atagtagtggaaaagaatlaaataattataaaattt 1709
Db 3760 gatagatgcagtgagaatcccatagaataatcttcagattcctaagaagaagactgcatc 3819
Qy 1710 aataagcaaaatagaataatgagatgcaaacattagtgtgaataatataaaagaatgaga 1769
Db 3820 tatlaaagataaagaagaagatgcttcactagtgtgtgaagaagttcaagacatgatat 3879
Qy 1770 acggttttggaaaaaatttacttaagacgaaataaacccagatgaaaaattttaga 1829
Db 3880 ggaatgaagatgttgagaagaatttgaattgaaaaataltggaagaggaagttatgaga 3939
Qy 1830 agtatctgacatgtlaaagttacaagttccaaaagtttataatgaacaaattgacga 1889
Db 3940 tgcgtgttgaat---aaatgacattactacgaacttatgagaagaactcaagattaa 3995
Qy 1890 attaaaaaagactccaattgtatttaaaaaatgtagaatlaaacaataatatactg--- 1945
Db 3996 atgaagtagaagcagatttaataaagatalggaaaaaatlaaagaattagaagaagcat 4055
Qy 1946 ttcccaattcttcaacaagaataatagaagaaccttattatgaattgtgtgaaa 2005
Db 4056 tatcagaagattcttaagaataatagatgcaaaaagatgatacatatgaaaaagttatg 4115
Qy 2006 aagaatltgataaaltanaagtgltcatgctlaaggtlagaatcatcatgataatgaga 2065
Db 4116 aagaaggaactgtataacgaagcgtgtgagtgaagttgtgagaatttaaaagatgctga 4175
Qy 2066 aaaaaaacataaagaagaagtgatcgtgatatccggaacatcaaccggaagggaaa 2125
Db 4176 aagaacaaagtcgaataaagaatctgcatlttaaaagatcttgaagaagatalatlaaa 4235
Qy 2126 taacagagaacatacaaacctcggaacaagaagactgctgttgaaaggagatt 2185
Db 4236 taaaagaataccaagaacttgaagtg---aaattttagaagattataaagaattaaaa 4292
Qy 2186 cagttacaagcacaagcacaagaacaaacaaagcacaacacacagttacccgttac 2245
Db 4293 ctattgaacagattatttagaagagaaaaagaataagaaagatcatttggaaaat 4352
Qy 2246 cagaagcaaaagcagacagtgcccaaccacacagcaccagttaaataaactggaatg 2305
Db 4353 tcgaagaagaagcctgagaataataaagatccttgaacagatatataataaagaagttact 4412
Qy 2306 ttcccaattagattacttgaaaaattatgaatt-----tttaatacttcatat 2358
Db 4413 cartagaagttgagaagaataaaatttagaagaagttacccgaattaaagaagagttag 4472
Qy 2359 atatgtcacaaatataatttggtttcacactcaactatgaaacgaaaagataltaaaacaa 2418
Db 4473 aacataataagtggtgtgctgcatalaaagtggttggaagaagtgatttgaagaag 4532
Qy 2419 tataaaattacaagaaggaaga--aagcaaatgaagttcatgtgaccttgacttat 2476
Db 4533 tagatgtattaaaggaatattatagacatgtttaaagggaataggaatttggagata 4592
Qy 2477 tgtttta-----tatacaataacatactcgttlaatgtattctatgtttagagcttaa 2530
Db 4593 tggataaggaagatttagaagatgtaacaaacacttggagaagaagttgaactcttaa 4652
Qy 2531 acaatagttatcacacacattatggaattatgaaaaagaagatggttgaattat 2590
Db 4653 aagaatgtttatctatgtcat---tagcatgtgataagaacaaatgaaacaagaaaaa 4709
Qy 2591 ataaacttaagataatgacaataatlaaaattatltagaggaagcgaaaaaagttacca 2650
Db 4710 aagctcaagaactaagtttggaagaagttatttaaaagaaggtttaaagaagaccaca 4769

Qy 2651 catctgtaaaaactctttcaagttccaaatgcaaacataltcatcattaaacactcaggata 2710
Db 4770 agaaaaaaataacaaaaaagaagaagttgatttgaatttaaggatlaaggaaaccaaaagtg 4829
Qy 2711 aaccggaagtaagtgcataatgatacatcacatcttcaaaatttgatlaattagtttaa 2770
Db 4830 aaatagtagaagttgaaatgaaagatgaaatataagagaagttgtagaagaagatatag 4889
Qy 2771 aattatttgaataacataatgagtccttggaaaaaaacaaaaatataatccagaattag 2830
Db 4890 aagaagatalagagaagaatlaaagttgaaagatalaagatgaagatalagatgaagatalag 4949
Qy 2831 gttcaaaaagtagtgaanaacttttlaagaaagatatlaaagaagtagtgcatactttata 2890
Db 4950 gttgaagcaagaatgagttatagatttaattagttccaaaagaagaacgcatggaaga 5009
Qy 2891 atgaatctttacaatttvtglaaaatctlaaagctgatgatataatcaatltgaatgag 2950
Db 5010 ttaagcgaaaaaagaataaatttagaaaaaagttgaagaagtc---gttagtgccttaa 5067
Qy 2951 aatcaaaaaggaagaatttagaagaagatatataaattaaaaaacttcaagttat 3010
Db 5068 aaacacgtagcaggaatgaaatgataatgttcaaaaatgtgaagaagttgataaga 5127
Qy 3011 cattgtattataataataataatlaaattagaagaagatttgaataaagaaga 3070
Db 5128 agtatctaaagctttgataatcaaaaaatgagttactaatgttttlaaacaatacaaga 5187
Qy 3071 --cagttgtaaatataaaltgcaaatlaaaaaacttacttataaagaacattaga 3128
Db 5188 tttttttagtaaaagttaaaacttcgtaaaaaatataaagatttgcgcacattcat 5247
Qy 3129 attcaaatggaattcacttaataaccgaagctgatatcaaaacttttctgtttctt 3188
Db 5248 atctgcgttgcagcattt---gcataatgtgagttggtgttcttcaatttctt 5301
Qy 3189 taacaaaaaaagaagcgtgaataatgagaagaacacacatgaaacacaaanaat 3248
Db 5302 attttcatcatgtgtatacaaatagcttcttcaacttcttatatccaaggttgacaac 5361
Qy 3249 attattgaaacattataaagagctgttlaaatalataatgtygaatcalctcattaa 3308
Db 5362 tataaataaataaagaagagagacgttttatttattcatttgaattga-----tatctttaa 5415
Qy 3309 aacttgaagtgagaatcaattcaaacagaagttatttgcagtttgaanaactttaa 3368
Db 5416 gaatttaaacattatttcaacaacatgaaagaataatttagtaagaagaataataaa 5475
Qy 3369 agtatt--aagtaatttagaagaagaatttaaggaatatttaatttagaagaagaanaa 3426
Db 5476 tgtatagaagttacaacaagaactgagaanaaagttatgtcaggtacagttacaataaac 5535
Qy 3427 ttatcacttacaagtgtgattacatcttaattgtcgaattlaaagaagaatlaanaa 3486
Db 5536 cgaagaacaacctaagttgataaaaaataaagttacgga--aaaagaagaagcga 5594
Qy 3487 aataaaaattatca 3501
Db 5595 aatcaaaaataaaaa 5609

RESULT 6
US-08-487-826B-13
Sequence 13, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim I.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.

OY	562	tgatgtyctaigtatcatctgltcaaaaccttccaattctaattctgtgcaatgaatt	641
Db	2442	tggtgctgtaaaaagctgycacaacaaattatbacaacacttttaagtaatttattgagtg	2501
OY	642	agacgtacttlaaaaactctglttcggtatagagaacatttagcaaatattaaagata	701
Db	2502	tatcgaaaactgtaggaataataagaacgtatattaaattgagatagagaagtaagaana	2561
OY	702	tgtaggaanaaayggaagatfacattaaaanaaalaanaaacatagaanaatatatga	761
Db	2562	tgtagtcaaccaactactagaanaacgtagagaacactacagctggaagtgtaactctt	2621
OY	762	attaattgagaagaagttagaanaacatctgttaaaataagaatgcaactaaagaaga	821
Db	2622	tagtacaataltagagagagacagaanaaatactattactatgatactatttagagaaa	2681
OY	822	aaaaaaaattatcaccaagctcaatgatacttctattcttacaataaacaattaga	881
Db	2682	attgagaagaaactccagaaaatgtaattaaagtgcgcgtttagaanaataccaaaagtga	2741
OY	882	agcacataattlaaagaagtttagaanaacglatgtgacatttaaaaaaaatgaaaa	941
Db	2742	ggaanaagagaagtagatagatgtaatttgagaagtagaagaaggtgcgtaccactt	2801
OY	942	cattaagaataattactgtatagataagataatgtaataatccccacgcgcgaattctg	1001
Db	2802	aatganaactgtggaacagcagagagaagaagagcgcgaataacatctacgaatatltga	2861
OY	1002	aatacaccaaatctctccctgtatagaacaanaaatcggagacacgaaagaagaat	1061
Db	2862	aaatttggagaanaatgcaatgagaagaatgaatgaanaatgttgcagaagaatttaga	2921
OY	1062	aaaaaattgcccanaacttlaattlaaata-----ctgtagtttatttctgaccc	1115
Db	2922	aaacgaacactgatttattactgtattgtagtaagtagaggaacacgtagaataatgag	2981
OY	1116	acttgaattgaataactcatttaagaagagaaaaataaataattgataagtgcagaagt	1175
Db	2982	agaaaggtttagaanaacatganaatgataagaagcattttttagtgaatatlttgaaagt	3041
OY	1176	tgaaacaaagaatcacact---gaaccaatgataatccaatgagatcttactcctt	1232
Db	3042	aaaaggaatcacagaaaatttataaaggtatglttcgaagtacagaacacgaatagt	3101
OY	1233	gtcataatagatlttaacatgctttaaagtaagtaacttacttctgttgatttaata	1292
Db	3102	aatccaatcagaagaagaaggttgatttgtaga---aatggtgtagtctcgatttga	3158
OY	1293	tcacattgtatatacaaaagaccgaatlaaaaaacatatatactgtataatgagaanaa	1352
Db	3159	taatatgaaaatltgaaaagaaggttatataaataattagaanaacttaccagactga	3218
OY	1353	atccataatgtaaatlttagagaaaaaattlaaataagaaaaaaaatttgatcgtata	1412
Db	3219	aggtgttccaagaacactgtaactgacaatgttagacaanaatgtatattgtagtgatgt	3278
OY	1413	aaaacttaccagacagacactgaagctttaaagtataacaacaaagaatatgaaat	1472
Db	3279	tcgcgtatagaagaatccaatttlttagaatataaaatgvgcgaagaaggttgaaagaat	3338
OY	1473	acttaatgaaattttagatgacaaattcaataataatagatlttaactaatlttoga	1532
Db	3339	gtttttaaatttggagaat---gtattaaagtggaagtgtgttaattactgtagaaga	3395
OY	1533	aatgatygttaaaagatatcatataaagtltgagaactttcacacacataactttgc	1592
Db	3396	aattagaatgaaacggttccaanaaggttagaanaaagaacgtttagtattatgaga	3455
OY	1593	atccatagaanaatcttcaacataatcttgaanaagttacaanaagctttaatata	1652
Db	3456	aatggaagaataattgttagatgtatlttagaggaagaanaagaagtttaacagaaga	3515
OY	1653	agattattctttaaaggaat---atagtagtgaanaaagaatttaaaatatltt	1709

Db	3516	gatagtgcagtagaagaatccatccatagaaatattcttcagatcttctaagaagaacatgc	3575
Oy	1710	aataagccaaatagaagaatagattggaacattagcttggaaaaatataaaaaagaagaa	1769
Db	3576	cattaaagataaagaaaaaaagatgttccactagctgtgttgaaagaagttcagaacatga	3635
Oy	1770	acagcttcttgaaaaaaaattactaagaacgaaaataaacagataagaaaaatttga	1829
Db	3636	ggatgaagaagctgtgaagaagcttctagatttgaaaaatactgaagaagggagttaatga	3695
Oy	1830	agttatcgacctgtgaaagtatacagttccaaaagtttatttaatgacacaaatttga	1889
Db	3696	tgctgttgaat---aaagacattactgcgaaccttctgaagaacatccaaagttaa	3751
Oy	1890	attaaaaaagcctaatctgtgttttlaaaaaatgtagatttlaaacgataatatacatg---	1945
Db	3752	atgaaagtgcgaacgatttataaagaatctggaaaatctaaagaatttgaagaagcat	3811
Oy	1946	ttcccaattcttaacaaacaaagaataagaacgaacctattatttaattgtgttga	2005
Db	3812	tatcagaagattcttaagaagaatataagatgcagaagatgatacatltagaagaagtattg	3871
Oy	2006	aagaaattgataaatttaaagtgcttcgcctaaagtagaatcatctgtataaatgtga	2065
Db	3872	aagaggaacatgatatatacgaacagcgtgtgataagttgttagaatttlaaagaatgcga	3931
Oy	2066	aaaaaacataaaaacgaagaagtcgaatcgatattctggaccatccaacggaagga	2125
Db	3932	aagacaaatctgaaaaaggtctcgtatttaaagatcttgaagaagatatataaaaga	3991
Oy	2126	taacaggaacgaacactacaacaccttgcacaacagcagatctgtttagaagga	2185
Db	3992	taaaagaatccaagaactctgaagtg---aaattttagaagattataaagaattaaaa	4048
Oy	2166	cagtaacgaagccaaagccaaagaacaaaacaaacagcacaccacgaatgccagta	2245
Db	4049	ctattgaaacagatattttagaagagaaaaagaataagaaaaagcatcttggaaat	4108
Oy	2246	cagaagccaaagccaaagctcccaaccacgaacagctaaataataaacttga	2305
Db	4109	tcgaagaagaagcgcgaagaataaagaatcttgaacgagatatataaagaagatact	4168
Oy	2306	ttccaattgatttactcttggaaaaatlatatattt-----tttaatactcatat	2358
Db	4169	cattgaagcttgaagaagaaaaaaatitgaagagtgatacgaattaaagaagggag	4228
Oy	2359	atagtccaaatatatttggcttccacactcaactatgaacgaaagatatlaaaca	2418
Db	4229	aacataatatagtgtagcgcataataaagtttgaagaagaatgatltgaagag	4288
Oy	2419	tataaaattacaagggggagag---aagcaatttaagttcagtgcattacattat	2476
Db	4289	tagatgatttaaaagggaatatacttcgacatgtttaaaggagatctgaaattgggga	4348
Oy	2477	tgattta-----tatcaaaataacatccctgtaatgtattctatgttttagacttaa	2530
Db	4349	tgatagaagaaagttctgaagatgtatacaaaaactctggagaagagttgaaccttaa	4408
Oy	2531	acaatagttataccaactattatygaaatttatgaaagaagaatgtgttgaattat	2590
Db	4409	aagtgctttatcttagtgactta---gcatgtgtagaagaacaaatgaaacaaga	4465
Oy	2591	ataaacttaagataatgacaaaaatataaatttatitgaggaagcgaaaaaagtata	2650
Db	4466	aagctcaaaagccttaagtgtgaaagagtattctlaaaagaagaggttlaaagaacaa	4525
Oy	2651	catctgtaaaacactcttaagttcctaattgaacacattatcatctaacacctga	2710
Db	4526	agaaaaataacaaaaaagaagtgataggttggatattaagataaggaaccaaagtg	4585
Oy	2711	aaccgaagtaagtcacaaatgatacattccatctcaaatcttgaatatagtttaa	2770

[illegible]

RESULT 9
 US-08-323-170B-1
 ; Sequence 1, Application US/08323170B
 ; Patent No. 5733772
 ;
 ; GENERAL INFORMATION:
 ;
 ; APPLICANT: Williamson, Kim C.
 ;
 ; APPLICANT: Kaslow, David C.
 ;
 ; TITLE OF INVENTION: Cloning and Expression of Plasmodium
 ;
 ; TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, Pf2300
 ;
 ; NUMBER OF SEQUENCES: 4

```

CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,170B
FILING DATE: 13-OCT-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/010,409
FILING DATE: 29-JAN-1993

ATTORNEY/AGENT INFORMATION:
NAME: Quine, Jonathan A.
REGISTRATION NUMBER: P-41,261
REFERENCE/DOCKET NUMBER: 015280-113100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO.: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 149..9556
US-08-323-170B-1

```

	Query Match	2.4%	Score 118.6	DB 1:	Length 9636;
	Best Local Similarity	42.2%;	Pred. No.7.2e-11;		
	Matches 1369;	Conservative 0;	Mismatches 1819;	Indels 59;	Gaps 10.
Oy	366	taaactcgaagctgatttaaacaacagatgaagaatacttcgttaacctcaagaact	425		
Db	4419	TATATCATATTACGTCTTTCCAAAAAGACGACAAGCTAAAGGTTTCATTTGATAA--	4476		
Oy	426	caaatccccaactctttgatlltaactaatcatatgtbaacttctgttgabaaatca	485		
Db	4477	TGAACATGCACACTATGTTTACTTATTAATTAACATATGTAATAAAAATTTGATTATATGATGC	4536		
Oy	486	tggttccaacttttaattlgtatgtagatgaagaattlaagtatatcatataaaattaa	545		
Db	4537	TAAACCTTAAAGAATTTGGATGATGGTTGCTTGCCTCGTAGCTACCTTAAATTAACCAATTG	4596		
Oy	546	ccttatcttgcattatttaagagcaaaaataatgatgtatgcttaagtattctgca	605		
Db	4597	TTTTAAAGATGCAATAGTACATACCAAATTTTAAACAATATTAAATGGTATACCTTATTATTA	4656		
Oy	606	aataccttcaaactttaaactcgycgaatgaatltgaagctacttaaaaaacttglyt	665		
Db	4657	AAATATATTTGGGTACTTTACATTATTAACATCAATATTATATATATGGAATATCCAGCTTT	4716		
Oy	666	cggatataagaaaccatttagacaatatlaagaataatgtaggaanaatrgaagaattacat	725		
Db	4717	AATGATATGATATATATCATTTAAATGTATATATGTTGATTTAAAAAATAAATAATATA	4776		
Oy	726	taaaaaaaataaaaaaacocatagaaaataitaaibaattlaattlgaagaagylgaagaaac	785		
Db	4777	TGTCAATTCACCATTAGGACCTAAAGCTTTTACGTCTCTTTATTAATAAAATTAATATATAA	4836		
Oy	786	aattataaaaalaaagatgcacctaagaagaagaaataaaaaataatataaccaagctca	845		

Db 4837 ATTGTAATTAATGTTACTGGCAGCTGATCAATAATAATCTTATGACATATATGATTT 4896
 QY 846 atatgatcttcttatttaacaataatagaagaacacataattataagcttt 905
 Db 4897 ACATTTATCTCAATTAACGTAATTAATTTAAAGAAATTTATTCATGATTTGGTAAAAAAA 4956
 QY 906 agaaaaacglatctgacttaaaaaaaatgaacatlaagaaatctactgataagat 965
 Db 4957 ACCACAGATTAACAGATGCTAACCCCTGAACTCTATATGCAATCTTTAAGTATTAATTAATC 5016
 QY 966 aaatgaatlaaanaatccccacggccaatctcggaatataccaatactctcctga 1025
 Db 5017 TAATGAATCTGCACTTTCTTCAACCGGAGATGTAGATGACAGAACTTTAATATTAAGAG 5076
 QY 1026 taagacaaaaaaatcgcgggaacacgaaaaaagaataaagaagaatcgccaatacttaa 1085
 Db 5077 AATGATATATGCGAAAGTTTATATCATGACAAATTAAGAAAGTATATATATATATATAT 5136
 QY 1086 atttaatatgaatglatltaactgactcactgcaatlagaatatactatlaagagaana 1145
 Db 5137 TGAATCTTTAGAAATTAAGATATTTGACAAATATGTTTACAAATTAATTTAAACCTCC 5196
 QY 1146 aaataaataaltgatalataagtcgaaggctgaacacaaagaaatcaactgaacccaatga 1205
 Db 5197 AAAATTAATGATGCTGCTCAATTCATATATATAGCATGTATGTATGCTTCACAAAAA 5256
 QY 1206 atatccaatlgagatctactatcccttgcatataacgataltaacaaatgctttaatga 1265
 Db 5257 TAATTTTAATTT---GTACCGAATCATTTAAAAAAAAGAGAGCTTGGTGTATATCCACT 5313
 QY 1266 acttaatctcttgatglatlaatlaatcacttgatlaacaaaagaacaaaglaaaaaa 1325
 Db 5314 AATATTTCTATGTTATGATATTTAAACCTTATAGATATATATATATATATATATATATAT 5373
 QY 1326 catatactgataatgaagaanaaaatcatalaaatgaatlaagaaanaaaltlaaanaat 1385
 Db 5374 ATCAAAAGATTAATATGAAAGCTGTAAAGCTAAACATATCTGAAACGACAAATGAAATNTGA 5433
 QY 1386 agaaaaaaaaaaatlgaaatcgtataaanaatcttaagagaagatcgaatctttaa 1445
 Db 5434 GTTACAAATTAATATCTTTATATGAAAAAAGATTTTCATATTTTGACACCTTTCGAATCGAA 5493
 QY 1446 gataatacaaaagaatlaaanaaaltlaactlaatgaatlaatlaagacaaatlaacataa 1505
 Db 5494 GAAACCTGGAATGAGATGTAGTACATATATGTTGTAGATA----- 5540
 QY 1506 taatatagatlaactaatctcgaaaaaagtatggtaaaagatalcatlaaagtga 1565
 Db 5541 -----CTGACCTGTATATGATTAACAGTACATTTGAAAAATAT- 5589
 QY 1566 gaactatacacacataactcttgacatccatagaanaatcttaacataacttgaana 1625
 Db 5590 AAAAATTAACACAGATTAATTTTGTGAGAAAGTTAATTAATGATATGATGATCTGAAGA 5649
 QY 1626 gttacaagaagctcttaaatatataatgaagaatlaacttlaagaaatlagatglaa 1685
 Db 5650 AAAAAAGATTTAGAAAGATATATACCTGGGCTATGTTAGTCCATGAAGTTTAA 5709
 QY 1686 agaatlaaanaatctlaaanaatcttaagaagaanaatagaatgagaatgaaacatlagt 1745
 Db 5710 AAAAAAGATCTTTTACATCATATGCTGCTTTGTTGTTCCACCAATTTGCCCAAGA 5769
 QY 1746 tgaanaatlaaanaagatgaagaacagcttctgaaaaaaaatcttaagaagaanaa 1805
 Db 5770 TTTTACATTTTAAGTAAGTAATTAATTAACAAATTAATTAAGTAAGTAATTAAG 5829
 QY 1806 taacacagaatgaanaaaat-cttagaaglatcgcacatglaaagatgaagcttaagaag 1864
 Db 5830 TCGATATATATGATATATATGATATGATATGATATGATATGATATGATATGATATGATATG 5889
 QY 1865 tcttataatgaacaaatgaagaatlaaanaagaatcgaatcttlaaanaatgtag 1924
 Db 5890 TGATTTCTCTAGAACAAATAGTTCTATTTTAAACATCCAGTGAATTAATTAATTAAGTAAG 5949

QY 1925 aatlaaacataatatacatgltcccaatcttacaacagaanaatlaagcaagaacct 1984
 Db 5950 AACTAAAAATTTGTGANA-----TAAATTAATTAATTAATTAATTAATTAATTAATTAAT 5994
 QY 1985 atttaataatctgltgaaaaaagaatctgaataatlaaanaatgltcagcctaagtag 2044
 Db 5995 TATCATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6054
 QY 2045 aatcaltgaataatgaagaanaaaacataaanaacagaagatcgaatcglaaattcgg 2104
 Db 6055 TAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6114
 QY 2105 aacacatacagaagaagaataatagaagaagaacacacacacacacacacacacacacacac 2164
 Db 6115 ATTAATTAATTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6174
 QY 2165 gatctgcttgaagaagagatcagltacagcagaacagaagaacagaacagaacacacacacac 2224
 Db 6175 TAATTAATTAATTTCTGTTTCAATGTAACATCAGTACTAATTAATTAATTAATTAATTAATTAAT 6234
 QY 2225 caccagttaccagttaccagttaccagttaccagttaccagttaccagttaccagttaccagtt 2284
 Db 6235 AGTTACTCTAATAA- AAGATGAAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 6293
 QY 2285 taataataaacaatcgaanaatgltcccaatctgaatctgaataatlaaanaatlaaanaat 2344
 Db 6294 AACATTAATTAATGATCTTTATATGATTAATTTCTGATATATCTGATTAATTAATTAATTAAT 6353
 QY 2345 taataatcatalatataatgacaaatataatcttgatcacaactcaactcaactcaactcaact 2404
 Db 6354 AATAAATTAATTTATGTAATAATGATGCAAGCTGTTCCAGAAATTAATTAATTAATTAATTAAT 6413
 QY 2405 agatataaacaataataaataatcaagaaggaagaagaagaagaagaagaagaagaagaagaaga 2464
 Db 6414 CAATTAATTTGGAACAAAGAAATGTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 6473
 QY 2465 catlagactatgltgltatatacaaaatatacaatctglaagatctglaagatctglaagatct 2524
 Db 6474 CATTAACCAATCATTAATTAATTTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6533
 QY 2525 gctlaaacaatgltatatacaactatlaatgaanaatlaatgaanaagaagaagaatgltgta 2584
 Db 6534 AAAAAATGAGATATCTATTAATTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAG 6593
 QY 2585 attataataactlaagaaatgaagaataatgaagaataatgaagaataatgaagaataatgaaga 2644
 Db 6594 CTTTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6653
 QY 2645 tatccacatctgaanaaactcttcaagttcatlaaagatgaagaacatatacatlaaagaacct 2704
 Db 6654 AAGATTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6713
 QY 2705 agatataaaccggaagtagtgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 2764
 Db 6714 GAGGATTAATTAATTAATTTTGTGAGAAATGATATCAAAATTAATTAATTAATTAATTAATTAAT 6773
 QY 2765 gttlaaataatctgaanaacatgagctctggaanaaacaacaaatatacatcaagaat 2824
 Db 6774 ATTTCAACACAGTTTCAGCTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6833
 QY 2825 taatggtcaaaaagtagtgaanaacttlaatgaagaatlaaagaatlaaagaatlaaagaat 2884
 Db 6834 GCACCTGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6893
 QY 2885 tctataatgaatcttlaaanaatctglaaanaatcgaagaatgaatgaatgaatgaatgaatgaatga 2944
 Db 6894 TATTTCCACACAGCTTTTATGTAATGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6953
 QY 2945 atgataatcaaaaagaagaattagaagaatataatlaaanaaacttaac 3004
 Db 6954 ATATGTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7013

1029 gaacacaaatcagagacacgaaaaagaatcaaaagaatctgccaacttaatt 1088
Db 2779 TTATTTATTCAGATTTTAAATATATATAGTATGATGCACTTTTATCTAATGCTCTGG 2838
Qy 1089 taactatgagttatcttaactgacccacttgaaatgaactatttaagaaaaaa 1148
Db 2839 AGATCTTAAAGATGCTGATTTCTTATTAAGAAATTTGCAAGATGTTTAAACAAATTA 2898
Qy 1149 taagaatctgataagtcgcaagcttgaaacaaagaaatcaacgaacccaatga --- 1205
Db 2899 AATTATATTTGGTATTTGTTTATAGTATACACTTATTAATATCATTTAGGTTTGACAC 2958
Qy 1206 ataccacaaatgagttactatcccttgatcataaagcatttaacaaatgctttaatga 1265
Db 2959 ATTTAAATGAATATGTTATGATAGTGTATTAATCAACCGTAAATCAATATATAGTATA 3018
Qy 1266 acttaactcttctggtgatttaattcaatcattgattalacaaagaacgaatga 1325
Db 3019 TATATGTTACATTTACCTCACAAATCATGATGATCTGTTTAAAGAAAAATCAATTTTAAA 3078
Qy 1326 catatatactgataagaagaaaaaatccataaagttaagaatgaagaaaaatga 1385
Db 3079 AGAAAAAGAGCTTGCATTAATGTTATTAATGCTTAATGATTAATCTATAGAGCTATTTTC 3138
Qy 1386 agaaaaaaatcgaacttgaaatctgataaaaaatcttaagaagacagatcgaacttaaa 1445
Db 3139 ACATAAATATGGAAGATTTTATAGTGTCCAGTTTCATCCGAGGATTAATATGCTCTGA 3198
Qy 1446 tgaataacacaaagaatgaagaaatcttaacttaagaaatctatgat- agcaaatcaata 1504
Db 3199 AGATATATCATTTTATTTTAAAGATTTTCTTTAGATATCTTTAAATGAAGAAAAACATA 3258
Qy 1505 ataatagatttaacttaattctgaagaaatgaagtgaaagaatcttaataagttg 1564
Db 3259 TACAGATATTTTGGATATTAATTTTATTAATTAAGAAAAAGAGCTTTTATAGTATG 3318
Qy 1565 agaacttacaacacaaatcttaactcttgacactgaagaaatcttaacataacttgaa 1624
Db 3319 TCGTGTATATGATATGAGCAACAGAGCAATTCGATTTATTCAGAAACACAAATTA 3378
Qy 1625 agttaacaaagctcttaataatataatgaagattatctttaaagaatataatgcttgaa 1684
Db 3379 AACTTTAAAGAAATGCTATATATATGTTATTAATTAAT- TAACTCTAACATGCAACGTTTC 3437
Qy 1685 aagaataaatacttaaaatcttaaaatgaagcaaaatgaagaaatgaagaaatctg 1744
Db 3438 AATCATCAAAAGCTTTGCGAGATAGATATGTTTACCACTTAATTTGGAATTTGTAG 3497
Qy 1745 ctgaataatcttaaaagaatgaagaaagctcttgaaaaaaatcttaacaaagcga 1804
Db 3498 AAAAAATTTTAAAGAAAGCAACCTGATTTTATTTATGATGATTTGGTGCAGACAG 3557
Qy 1805 ataaacagatgaaaaaatcttgaaagatctgacattgtaaaagtaacaaag 1864
Db 3558 CTTTAAATTTGCTTTTAAATGTTAGATCAAAAAAGATTTGAAAGAAATTTGTCAT 3617
Qy 1865 ttctataatgaacaaatctgacgaatttaaaagctcaatctgatttaaaaaatgtag 1924
Db 3618 GTTTAGTACATCTTTAGATCTATTAAGATTAACAGAAATTAACATTTTGTGCGAAA 3677
Qy 1925 aattaaacataatataatctgctccaatcttaacaaacagaanaaagcaagaaact 1984
Db 3678 AATTAAAGAAATTTATGAAGAAATGCTCCATATGTTAGTCAAAAAATGTTAATCAG 3737
Qy 1985 attatttaattgcttgaaaaagaagaatctgataaatttaaaagtctcaagcgaagtag 2044
Db 3738 CTATTTATATATGCTTAATTAATTAAGATGATTCATATTAATGATACATTTTCTGTAG 3797
Qy 2045 aatcatctgataaagtaagaaaaaaaacataaaacagaa --- gctcaatcgtaaat 2101
Db 3798 GAGGATTAATATGATCTTTTATTAATTAATGAAGAGAACTTATCGAAAAATGTTAATTA 3857
Qy 2102 cgaacacataacaggaagagaataacaggaacgaacactaacaacactgacacag 2161

Db 3858 TATTTTACAACTGATTAATGAATATTTTATATATTAATATTAATTAACAAAGATGGAAGAAA 3917
Qy 2162 caggaatcgtcttaagaagagaltcagtlacaagcacaaagacaaagaacaaacagcag 2221
Db 3918 TAGAATATGAATTTATTAAGGATTAATTAATTAATTTGTTATAGCTTAATTAATGGA 3977
Qy 2222 aaccacagtagcagtagccagtagcagaaagcaagacaaagctccacacacagcagc 2281
Db 3978 ATATAGATCCATTTAGGTATATCATACAGAGATATGTTATGTTGTTGCCACTTCAAAACAT 4037
Qy 2282 cagtaataaataaacttgaaatctgtccaaatgaattatcttgaaaaatlatgaat 2341
Db 4038 TAAGTATATTAATTAATTTAATTAATTAAGAAATATAGCAATATAGCAT --- TAAAGTATTAAC 4091
Qy 2342 tttaaatcctatataatctgacaaatataatcttggtttcaacactcaatgaag 2401
Db 4092 ATTTAAATATTAATGAGAGATTAATTAATTAATCA -ATTTGATTAATTAATCCAAACAGAGAA 4150
Qy 2402 aagaatlataaacaaatlaaaatlaacaaaggaagaaagcaaatlaagtlcgtg 2461
Db 4151 TATTTATTTATTAAGTTAATGCTTAGGCTTAGAGATTGACCATTAATGCTTTAAAGCT 4210
Qy 2462 atccatlaagactatctgttlaaataataaataacacacacacacacacacacacac 2521
Db 4211 ACTGTTATCCACTGCTTAT 4265
Qy 2522 atagcttaacaaatagttatcaacacattatgaaatcttgaaagaaagaaagtggt 2581
Db 4266 TAAATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4325
Qy 2582 gtaattataaacttaagaagaaatgaacaaatlaaaatcttaagaaagaaagaa 2641
Db 4326 ACATTAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4385
Qy 2642 aaglatcacaatctgtaaaactcttcaagtlcaatcaatgaacacatlaacac 2701
Db 4386 TGAATAGTATGAAGAAAGTGTAGGAGAAATGATGCTATAGCTTAAGCAACCTTGAAGAAAT 4445
Qy 2702 ctcaagataaacccgaagaaagtaagtaagtaagtaagtaagtaagtaagtaagtaag 2760
Db 4446 CTATACAAATTTTAAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4505
Qy 2761 -aatagtttaaacttcttgaaacacatgagctcttgaaagaaacaaatataacaa 2819
Db 4506 GTATAGATTCGATGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4565
Qy 2820 agaatla -atagtcacaaagaaagtagtaaaactcttgaaagaaatataaagat -agt 2877
Db 4566 TTGATGCTATATATCAAGCTTTTCCATTTTAAATTTGCTTATGATTAATTAATTAATTAAT 4625
Qy 2878 gatacttataatgaatcttatacaaatcttgaaatcttaagctgaatgaatgaat 2937
Db 4626 CACATATATTAATTTGCTTTATCAATTAATTTTATTAATTAATTAATTAATTAATTAAT 4685
Qy 2938 tcaatgaatgaatcaacaaagaaagaaatgaagaatataatlaaataaataa 2997
Db 4686 TGAAGAGTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4745
Qy 2998 acttaacagttatcaatctgattataatataaataaataaataaataaataaataa 3057
Db 4746 GTTTTATGATTAAGCAAAATAGCTCACTACTATTCCTTCAACACAGCAATTAATTAATTA 4805
Qy 3058 gataaaagaaacagcttgtaataatataaagcaaatlaaaactcttataa 3117
Db 4806 AT 4865
Qy 3118 gaaca -attagaaacaaatgaatcttaataaataaataaataaataaataaataa 3175
Db 4866 AGCTAGATTTATTTCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4925
Qy 3176 ttctgttcttcttaacaaagaaagaaagctgaagaaatgaacaaactgaacacactg 3235

Db 4926 TAACTAATTTATTTATTTAACTTCAAGTCAGACATGATGTTCCCATTAATA 4985
OY 3236 aaaaacaaaatattattgaaacattaaagacttgtaaatatttaattgagaa 3295
Db 4986 TGAAGAAGGAAAAAGATATGCACCTTAATATATTAAGCAAAATGCAAAATAGA 5036
OY 3296 catccattaaaacttaagtgaaagaaatcaatcaacagagaataattatgcagtt 3355
Db 5037 -----AAAAGTCATGTCAGAACACACTTATATATGAAGTAGTTGATGAAG 5086
OY 3356 tgaagaactttaagatlaagtaaattagaagaaataaaggataatttaattag 3415
Db 5087 GATACCAATTTACACAAAGAAATTAATTAATTAATTAATTAATTAATTAATTA 5144
OY 3416 aaaaagaa---aaattatcatcttaacgtgattacatcttaattgcgaaatla 3472
Db 5145 AAAATTAATGTAATTTGAATTAAGAAATCTATGCTATATTAATTTCTTCAATGTA 5204
OY 3473 aagaaglaaataaaaataaaattatagagtaattctccaaagtgaaataatagatg 3532
Db 5205 ATACAAATATATTAATTAATGAAGAAATTAATTTGTCATGATATATCTATTAACAAAATA 5264
OY 3533 ttaacaatgcatagaaatcttaacaaaattctcccaagaaagacagatgtgcacag 3592
Db 5265 TAAAGCTTCAATTAACCAATTCCAATTAATTTATATGCAATTAATGAAGAAATGTTGAACAA 5324
OY 3593 ttgtaagtgaaagtgatgcacacatctagaaagaagtaaccagaagaacccgcacaa 3652
Db 5325 ACTTAA---ATTGTTATCTGAAAGGGCGGTAGCCATCATATATTTGTTAAAGAAAGAA 5381
OY 3653 ctcatgtagagcagagcttcaacacaaatcaacaaatgctcagatgaagatg-- 3710
Db 5382 AGAGTATAGATCTGATGATACCAATATTTTAAAGTGCACAAATTCGAATTAATTAATCTTTT 5441
OY 3711 -----agatgacgtatcatagtaactatattgagagaatccgaagaagatgatg 3763
Db 5442 CATGTATATGATGAATTAATGAATTAAGCAACGTTGATTAATGATTAATGAAGAAATGATA 5501
OY 3764 atttagacaagtagtaacagagagagcagtaactctccgaatctgtaacatactt 3823
Db 5502 CGAAAAACGACAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5561
OY 3824 cttaaaattgaaatgaaatgaggttttaatttaaaaccttaagcaggtgtttataga 3883
Db 5562 TTAATTAATTAAGATTAAGAAAGATGTTCTTATTAAGAAAGTGAAGAAAGATTTGTGA 5621
OY 3884 gttaaaaaaacaaatagaaataacgtttagacatttaagttaagttaaggaatctt 3943
Db 5622 TGGTAAATTTAAAGAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5681
OY 3944 taattcacgattttaaagcgtgaaatcttaaaatgttttaagaatcaagtttaattc 4003
Db 5682 AAAAGATTTGCGTGAAGAAAGAAATTAATTAATGAAGAAATGTA---AACATTAAGAAAGAA 5738
OY 4004 catataaagatttaacatcaagtaattatgtgttcaagaagacataataattcttaata 4063
Db 5739 AATTAATTAAGATTAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5798
OY 4064 aagaaaaaagagataaattcttaagcagttataattataatgaattcaatgaatgag 4123
Db 5799 ATTCAACTAATGATCATTTTATTTATTTAGATTAATTTTATACATCAGATGAAGAAATAGCGG 5858
OY 4124 atataaatttgcaaatgagttcttgatlatataaaatattatccgaaataataat 4183
Db 5859 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5918
OY 4184 cagatttagattcaatttaaaaaatataatcaacagacaagaagtgaaaaatgaaatc 4243
Db 5919 GTAATTTCTTATTAATTTAGATTTCCGTAATTAATTAATTAATTAATTAATTAATTAATTA 5975
OY 4244 ttcccttttaacaaatattgagaccttatataaacaagtttaagttaaatgagatg--tta 4302
Db 5976 TGAAGAGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6035

OY 4303 ttgttaattttagaagcaaaagttctaaattatatacatatagagaatcaacgttagaa 4362
Db 6036 ATTGCTCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6095
OY 4363 gttaaaaataaagaacttaatttacttaaaaaacaattcaagaacaattgycagatttaa 4422
Db 6096 TGAAGAAATGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6155
OY 4423 aaaa 4426
Db 6156 ACAA 6159

RESULT 12
US-09-150-741-1
; Sequence 1, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
; TITLE OF INVENTION: Synthetase II
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/150,741
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: A093/00617
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446,855
; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 8920
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
US-09-150-741-1

Query Match 2.1%; Score 102; DB 4; Length 8920;
Best Local Similarity 42.7%; Pred. No. 2.9e-08;
Matches 1693; Conservative 0; Mismatches 2200; Indels 71; Gaps 20;
OY 489 ttcaaatatttaattgagagataatgaagaatlaattgaattatataataaactt 548
Db 2241 ttcaaatatttaattgagagataatgaagaatlaattgaattatataataaactt 2300
OY 549 ttatttgaatttaagagcaaaattaaatgtaatgtaatgtaatgtaatgtaatgtaatg 608
Db 2301 gtaatgtaatgtaatgtaatgtaatgtaatgtaatgtaatgtaatgtaatgtaatg 2358
OY 609 accattcaatttaaaatttgcgaatgaattagacgttaacttaaaacttgcgtcg 668
Db 2359 taataatgtaatgtaatgtaatgtaatgtaatgtaatgtaatgtaatgtaatgtaatg 2418
OY 669 atatagaaaacattagacaatttaagaataatgtagaagaatgtagaagaatgtaga 728
Db 2419 taataaatgagaacaaacaaacaaatagcaaaataacaaataacaaataacaaata 2478
OY 729 aaaaaataaaacaaacaaacaaataaaatgaatttaattgaagaagaatgaagaacaat 788
Db 2479 taacgaggttgattatataaagaagaatgaagaataatgtaatgtaatgtaatgtaatg 2538
OY 789 tgaataaataaagaatgcaacttaagaagaagaagaagaagaagaagaagaagaagaaga 848
Db 2539 tagcaatataatataatgacacaaataatagacataccgaatttaatttaaatgaatga 2598
OY 849 tgaattcttatttaacaaataaacaataatgaagaagaagaagaagaagaagaagaaga 908
Db 2599 ttatttacttattatagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2658

QY	909	aaagcgtttgacactttaaaaaaaatgaaacataaaggaattacttgataagaataa	968
Db	2659	aagaaaagtagacccaataaagaaaaaatctgttaattgttgattggtatgataaaatag	2718
QY	969	tgaatttaaatccccccacgcgcacattctgtgaataacccaataactcttcgtgata	1028
Db	2719	tataatcaaaatttaataagaacggtagatcttcctaacaatacatattatgtaac	2776
QY	1029	gaacaaaaaatcgagacacgcagaaaagaataaagaatttgcaaaacttaaat	1088
Db	2779	ttattattacaattttaaccatacatagtttagtgacgctcttttataatggtccg	2838
QY	1089	taatttgaatggttatttactgtaccattgaaatagaatactattagaagaaaaa	1148
Db	2839	agatccttaaaaagtgtagtttcctttaaaaaatttgaagaatagttttaacaaaata	2898
QY	1149	taaaatatgtataaagtgcaaaagcttggaacaaaggaatcaactgaaccaatga---	1209
Db	2899	aattattcttgtaattgttaaggtaacaaacttaagtalactatctgtgtgagac	2958
QY	1206	atatacaaatggaattacttaacctgtgataaagaatataaacaatgcuttaatga	1266
Db	2959	atataaaatggaataatgtaataaggtggttaatacaaccgtaatacaattgataa	3018
QY	1266	acttaattccttggtgatttaatttaaccattgattatacaagaaccaagtaaaa	1326
Db	3019	tatagttacaatccctacaacaaatcagtactggtttaagaaaaaacttaattaa	3078
QY	1326	catataactgataatgaaaagaaaaaatccataaataaggaattagaagaaaattaaat	1388
Db	3079	aagaaaagaagcttcgattagttataataatgcataagataaaactataagaagatttc	3138
QY	1386	agaaaaaagaaaatctgaatctgataaaaaacttcaagaaacagaacctaatctttaa	1445
Db	3139	acataaaaatggaagatttttagtgycacagtttaccgcgagggaataatggtccgga	3198
QY	1446	tgataaacaaaaggaatatgaaaaaattacttaatgaattatgat-agcaaatcaata	1504
Db	3199	agatacaccatttttaatttaagaattctctttagatatactttaagaaaaaacaata	3258
QY	1505	ataaatagatttaacttaattcgaaaaatlgatgtaagaatatctataaagt	1566
Db	3259	taggaataatlttagabataataatattataataaaaaagaaagtgcttcttttagtag	3318
QY	1565	agaaactcacacacataatacttttcacacctagaataatctcaacaatacttgaa	1624
Db	3319	tgggtgtttagttagtagacagagagagatccgattatccagacaccaaagcaatca	3378
QY	1625	agttacaacaaagctctaaatatatgtagaatatctctttaaaggaatagtagtgaa	1684
Db	3379	aggtttaaagaagtgtagatatactgtatacttagt-taacctcaacagaaactgttc	3438
QY	1685	aagaaatthaattatataaaaaatttaataagcaaaatagaataatggaattgaacattg	1744
Db	3438	aaacttcaaaaggttgcgcagataaggtlaactctttcacagtttaattgtaatttgag	3497
QY	1745	ttgaaatatttaaaaaagatgaaagacagcttttgaaaaaaaatttactaaagcgaa	1804
Db	3498	aaaaaatataaaaaagaaaaaacctgatttttttagtacaattggtggtcgacag	3557
QY	1805	ataaacacagatgaaaaaattttagaagtalctgacatgttaaagtaacagttcaaaaag	1864
Db	3558	ctttaaatggtctttaaigttagatcaaaaaaagtatgtaaaaaagaataattgcaat	3617
QY	1865	ttttattatgaaacaaaattgacgaattaaaaaagctcaattgattttaaanaattgag	1924
Db	3618	gttttagtacaattcttagaactataaagaataacgaaataagaaactattttgttgaa	3677
QY	1925	aatttaaacataatatacatagttcccaactcttcaacaaacagaanaataagcaagaact	1984
Db	3678	aatttaaaagaaattatgaaagaatagctccataatgtagtgtagtgaaaaaatgttaataag	3737
QY	1985	attatttaattggttgaaaaaagaaattgataaatttaaaaggttcaatgacctaaagtag	2044

Db	3738	ctattgatatagctaatanaaataaggatatccaaatattagtagtacaacatttgcgttag	2101
Qy	2045	aatctcttatataatgaaganaaanaaacataaagaagaa---ggttcattgcgttaatt	2101
Db	3798	gaaggatcaatagtagtttccataataatgaagaagaacttaccganaaaytataaata	3857
Qy	2102	cgaacccaataccgaagaganaatacaagacaagacaatacaaaaccttgcacaaga	2161
Db	3858	tatttttaacaactgatatgaatatttataagataatcatttacaaggaatggaagaata	3917
Qy	2162	caagatctgccttttgaagaagagattcagtagcaagacaagacaagaacaaaacaagac	2221
Db	3918	tagaatatgaaatttcaagagataatnaaataatttctatagctatattglaaatgyna	3977
Qy	2222	aaaccaagctccggtaccagtagccagaagacaagaagatcccaacaacacagacac	2281
Db	3978	atatagatccatttggtataatacaggaagatgtagttagttagttagcatttacaacat	4037
Qy	2282	ccgataaataaataacgnaaatttctccaacttagatttattcttgaanaatatatgat	2341
Db	4038	taagtaattatgaatatataaatttagagaataagcat-----taagyaatttaac	4091
Qy	2342	ttttaaacttcataatatalgccaataattttggttccactccaactatgaacg	2401
Db	4092	atttaaatattataggaagaatgtaatatca-atttgcgtataatccacaacagaga	4150
Qy	2402	aaaagaattaaacaatatanaaattcaagaagagagaagaagaagaatgaattgaatg	2461
Db	4151	tattgtattatttgaatgaatagtcaggcttagtagagaagttccagcatltagcttcaaac	4210
Qy	2462	atccattagacttattgttataatacaaaatacacaactcgttaattgtaattcattgttg	2521
Db	4211	actggttaccacttgcatttatataatcagaanaaataagctcttggatgatg----ttga	4265
Qy	2522	atagcttaacaatagtttatcacacatttatbgnaatttatcganaaagaatggttt	2581
Db	4266	taagtttaaaaaatagcatactaaaanaacaactcctggttttgaccctctctagatt	4325
Qy	2582	gnaatttataaacttaagaagatatgacaanaattnaaatttattagagaagaganaa	2641
Db	4326	acattacaacaanaataccacagatgggaatttaataaatttgaatttgcgttctaataca	4385
Qy	2642	aagttccacactctgtbaaaaactcttccaagtttccaaatgacaatgacattatcataaac	2701
Db	4386	tgaatagtagtatagaanaaagtgtagaagattatgcttatagtgtagaactttgaagat	4445
Qy	2702	ctcaggaataaacccgaagtgaagtgcgaatgatatcatcacatttcaaaatttgaat-	2760
Db	4446	ctatacaaaaactttagatgtagttgatataattatttagattttagtaatacgtatt	4505
Qy	2761	-aatggtttaaattatttgaanaacatatgagctcttgaanaaanaaanaatatatacca	2819
Db	4506	gtagatttggatggaagaagaataatttgaagaatttlaaanaatccatcccaaaaagaa	4565
Qy	2820	agaatta-atgggtcaaaaagtagtgaanaactttatgaaaagatatnaaagat-agt	2877
Db	4566	ttagatgcataatcaacgaagcttccattttaaataatgctcttgataaataacatgagctga	4625
Qy	2878	gatacatatttataatgaacttttacaanaatttgtlaaaatcttcaaaagctgatatat	2937
Db	4626	caaatattgatttgcgttcttacaataaatttataataataataatcttacaataagat	4685
Qy	2938	tcaattgaattgttgaatcaaaaaggaagaattagaagaagatatataaataaataa	2997
Db	4686	tgaanaacgttlaaaatttagagaacattcttctttaaattttagaagatttlaagaagctg	4745
Qy	2998	acttacaagttatcatattggtttatatataaataaataaattnaaatttgaagaattatt	3057
Db	4746	gttttagtgaatgaagcaaatagctccactacttacccttcaacaagcgatatataata	4805
Qy	3058	gataaagaanaacagtttgttaataataaattgcaatttanaaaacttacttttaaa	3117

Db 4806 ataataaataatgactgacgaggtacagaaaatgactgataaataatagagaa 4865
QY 3118 gaaca--attagaataaattgaattcaacttaataccaaagcatgattcaaaact 3175
Db 4866 agctaggaattattccacacataaagtataatgatacttaacgacgcgaattccggtc 4925
QY 3176 ttctcttcttaacaaaagaagctgaatagcagaactgaaacacatag 3235
Db 4926 taactaattattatttaactatcaagtcgaagacgtgattctccattaaata 4985
QY 3236 aaaaacaaaataattatgaaacattataaaggactgtttaattataatggtgaat 3295
Db 4986 tgaagaagaaaagatgacgcttaataataacgaaatgcaataaga----- 5036
QY 3296 catctcattaaaactttaaggaagaaatccaataagaagaataatgacggt 3355
Db 5037 -----aaaagctccgtcgaagaaccattataatgaagaagttgataag 5086
QY 3356 tagaaactttaagatttaagtaaatgaagaagaataatgaagaatttaattag 3415
Db 5087 gatacacaattacacaaagaataataataataataat--atgaattcgtgaatgtag 5144
QY 3416 aaaaaga--aaataatcattatcaagtgatatacatatttaattgctgaata 3472
Db 5145 aaataaatgttaaatgaaataagaaatccatgctataataattcttaattgataca 5204
QY 3473 aagaagtaataaaaataaataatatacaggttaattctccaagtgaataatagcagt 3532
Db 5205 atcaataataatataatagaataaataattgctatgataatactataacaaaata 5264
QY 3533 ttaacaatgcatatagatcttaacaaaattctccgaaggaagacagatgtgcacag 3592
Db 5265 taaaagttcaataaacaatccaattatctataatcgaaataatgaaatggtgaacaa 5324
QY 3593 ttgtaagtgaagtgatccgacacattagacaagaatcaaccaagaacccagatcaa 3652
Db 5325 acttaa--atgtgatactgaaaggccgtgacatcataatataatgtaaaagaa 5381
QY 3653 ctcatgtagagagagatctcaacaaataacacaaatcaaatgacgtgagagt-- 3710
Db 5382 agagttagagatctgataacaaatattttaagtcacaaaattcaaaataaacttt 5441
QY 3711 -----agatgacgtatcatagacttaatttgaagatccgaagaagattatg 3763
Db 5442 catgtaataatgagataatgaataaagcaagctgattgtaatgtaactgaataatgata 5501
QY 3764 attagagacagatgataacaggaagcagtaactcccttcgttaattgataacatactt 3823
Db 5502 cgaaaaaacgagagataataatactacaacagtaattatggaaggtaaaaatagtgta 5561
QY 3824 ctaaaattgaaaatgatatgaggttttaatttaaaacctttaggaaggtttatagaa 3883
Db 5562 ttaataataagaataaagaataagttcttaattgaaagtgatgaagaagatcatgga 5621
QY 3884 gtttaaaaaaacaattagaataaatacgtatagacatttaattgtaattgtaagatatt 3943
Db 5622 tggtaatttaaaaaagaataaataatataatgtaattgaataatgtaattgtaagga 5681
QY 3944 taaatccagatlttaaaagtgtaaaatccaataatggttttagaatacagatttaattc 4003
Db 5682 aaaaagataatggaataaaaaataataatgataatgata--aacataaagaataaata 5738
QY 4004 cataataagatttaacataagaataatgattgtaagaagtcacataataattcttaata 4063
Db 5739 aataataaagataatgagataataataataatagtagtgattatccaatggaacatcac 5798
QY 4064 aagaaaaaagagaataattcttaagcagttataatataatgaattcaatagatcag 4123
Db 5799 attcaactaagatcatattatattagataatttaataacatcagatgaagaataatgagga 5858
QY 4124 atataaatttgcaaatgagttcttgatataataataataatataatcgaataatataat 4183
Db 5859 ataataaaaatgataatgattattcttaagaagaaaaaagataataataaaccctg 5918

QY 4184 cagatttagattcaattaaaaatataatacagcaacaagaagtgaaatagaaatacc 4243
Db 5919 gtaattcttaattgtgtgattatcggtatataataatgataaataat--ataaga 5975
QY 4244 ttcccttttaacaatattgagacctataataaacaagttaaatgaattgatt- 4302
Db 5976 tgaagaagtttaataagaaatlttaaatgataatataataatgtaattgata 6035
QY 4303 ttgtaattcatttagaagcaaaagtcttaaatataacataatgagaatcaacgtagaa 4362
Db 6036 atggttaattataatgaatgtagtgcattgttaagtgaagagatagaatgataat 6095
QY 4363 gttaaaaataagaacttaattcttaaaaaacaattcaagaacaattgagcagatttaa 4422
Db 6096 tagaaatgattgataaaaaaataatgatacatatacaacaatataatcgtttaa 6155
QY 4423 aaaa 4426
Db 6156 acaa 6159

RESULT 13
US-07-792-865D-1
; Sequence 1, Application US/07792865D
; Patent No. 5646247
; GENERAL INFORMATION:
; APPLICANT: John W. Barnwell, Mary W. Galinski,
; APPLICANT: Samuel P. Wertheimer
; TITLE OF INVENTION: MEROZOITE ANTIGENS LOCALIZED AT
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Darby & Darby P.C.
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM or IBM-compatible
; OPERATING SYSTEM: PC/MS-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/792,865D
; FILING DATE: 19911004
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US90/01849
; FILING DATE: April 3, 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson, Arthur
; REGISTRATION NUMBER: 34,354
; REFERENCE/DOCKET NUMBER: 5986/14692-US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 527-7700
; TELEFAX: (212) 753-6237
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3763 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; DESCRIPTION:
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: P.vivax
; STRAIN:
; INDIVIDUAL ISOLATE:

DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY: Lambda gt 11 native P.vivax
LIBRARY: DNA expression library
CLONE: 5.3
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: This sequence corresponds to
OTHER INFORMATION: Figure 1A (sheets 1-4) in the
OTHER INFORMATION: Application, as filed.
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-07-792-865D-1

Query Match 2.18; Score 101.4; DB 1; Length 3763;
Best Local Similarity 43.3%; Pred. No. 3.3e-08;
Matches 665; Conservative 0; Mismatches 841; Indels 29; Gaps 3;

QY 409 ttaactcaagaacccaatataccccaactcttgatttaactataatgtaact 468
DB 1 TTGGATTAAGAAAAGTTAAAGATACAAAGTTTGATGAAAAAGAAAGCAATGAAAA 60
QY 469 ttgtgataatacttcgttccaatatttaattgatatgataagaatgaatga 528
DB 61 GCTTATGAAAAAATGGGAATACGCTTAAAGATTAGAAAAAATGGATGCGAAAAAAC 120
QY 529 ttattataaattaaactttatttgatttataagagcaaatataatgattgt 588
DB 121 ATGAAAAAGAAAGTAGAAGAGCTCAAAATACAAATACAAAAGAAATTTTATTGATCATGAT 180
QY 589 gctaagattattgtcaataacttcaatccttaaaattcgtagcaaatgatacgta 648
DB 181 GTTAAATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 649 cttaaaaaacttggttcgatatagaaaaacattagacatatataagaataatg 708
DB 241 TATTAATAAAGAAATGAGCAAAATTAAGAAAAAGAAAGCAATGATTAAGCAAGGTGATACA 300
QY 709 aaaaatggaagattcaattaaaaaataaaaaacattagaaaaatataatgaat 768
DB 301 TCTAAATTTTATTTATATACAGAACATACAACTGCTACAGAGTAAGCTTAATAATGAA 360
QY 769 gaagaaagttagaacaatttgataaaaaataagaatgcaactaagaagaagaaaaa 828
DB 361 CAATTTATTTATTTGCTTACGACAAAAAGAAAGCTGACACAAAGCCAGATATTAAC 420
QY 829 aaattatcaagctcaatatgattcttattcaataaacaattagaagaagacat 888
DB 421 GATTTGAAAGCAATTAAGAAAGAGGTGCAATTAATAATTTACAACTAGTCAACAAAGAAAGT 480
QY 889 aattataagcgttttagaaaaacglatgacactttaaaaaa-----aaatgaaac 942

DB 481 AATTCTATGAGGAAATCGCAAAAACAAATTCATGATGAGAGATTTCATAATTTGAAAC 540
QY 943 attaaagaattacttgataagaatgaatgaatgaatgaatgaatgaatgaatgaatga 1002
DB 541 AATTCGAAACTATGCTTAAGAAATATCAAAATTAATCAAAAGCAATTAAGTTTAAAG 600
QY 1003 aatacacaatactctcttgataagaacaaaaaaatcgaggaacacgaagaaaaata 1062
DB 601 GAGAAATGCAAAAACAAACTTAATTAAGAAAGCAATTAATGCAAGAGTGGCAGCATATG 660
QY 1063 aaagaatgccaaacttaatttaattgattgatttatttactgacactg-- 1120
DB 661 ATGAAAGAGCAAAAGCCACATTAAGCAATTTGACATACCTTTAGAAATGACCAATA 720
QY 1121 -----aattagatactatttaagaagaaaaataaaatattgata 1164
DB 721 GATACGAGGTAGCAAAATTTGAACAAATTAATGCTGAATTTATGAATTAAGAAAGATGAA 780
QY 1165 agtgcagaagttgaacaaaggaatcaactgaacccaatgaatatccaaatggaattact 1224
DB 781 ATTAATTCCTATTTTAAGTGAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG 840
QY 1225 tatcctttgataataagaatatacaatgacttaaatgaacttaattcttttgat 1284
DB 841 AATTCAAAAAGAGAAAGAAAGATTAATGAGTCTTGCAAAAATTTAAGCTTAAGAGAA 900
QY 1285 ttaattatccatttgattatatacaaaagcaagtaaaacatatatactgataatga 1344
DB 901 AGCAATTCGAATTAAGGTTAACTTAATTAAGAAATTAAGAAATTAAGAAATTCGAAACA 960
QY 1345 agaaaaaatctcaataatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 1404
DB 961 TACTTAAAGATTAAGAGAGCGAGAAAGCAAGCTTAAGTCAAAAGTGAATTAATTCAT 1020
QY 1405 tctgataaaaaactttagaagagacagatcgaatcctttaaattgataaaaaagaat 1464
DB 1021 AAACATGAACCAACTATACATTAATTTTCAAGCAATTCGAATTTTGAAGTGAAGCAAGT 1080
QY 1465 gaaaaattacttaataagaattttagaagaatctcaatataatataatgaatgaat 1524
DB 1081 AATTCGCAAAAATTAATTAAGCAAGAGACATTAATGAAGAAATTAAGCAATTAAGCA 1140
QY 1525 ttcgaaaaatgtaggtaaaagatatcaaaagttagaacttaccacccaat 1584
DB 1141 TCTGAATTTGAACACAGGTGAAGGTTTCCAGAAAAATTAATTAATTAATTAATTAATTA 1200
QY 1585 actttgcatcctatgaanaattctaaacatacttgaagaagttaaagaagctttaaa 1644
DB 1201 CATTAATTTATGCAACGACAGATGAATTAATTAATTAATTAATTAATTAATTAATTA 1260
QY 1645 tataatgaagattatctttaaagaatataatgattgtaaaaaagaatataatataa 1704
DB 1261 CTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1320
QY 1705 aatttaataagcaaaatagaanaatgagattgaacattgattgaatataatgaagaat 1764
DB 1321 CAGGAGGAGAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1380
QY 1765 gaagacagcttttgaaaaaaaatctaaagacgaatataaacagatgaagaaat 1824
DB 1381 TCAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1435
QY 1825 ttagaagttatgacatttgaagaatgaagtaagtaagtaagtaagtaagtaagtaag 1884
DB 1436 ATGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1495
QY 1885 gacgaatataaaagactcaatgatttataaaaaa 1919
DB 1496 GACTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1530

RESULT 14

